Lehrstuhl für
Methodik der Fernerkundung
Remote Sensing Technology

Technische Universität München
Ingenieurfakultät Bau Geo Umwelt
Lehrstuhl für Methodik der Fernerkundung

# Bayesian Inference, Applications in Persistent Scatterer Interferometric Synthetic Aperture Radar

**Homa Ansari**

## Master's Thesis

Earth Oriented Space Science and Technology – *ESPACE*

**Period:** April 15, 2013 – October 15, 2013

**Supervisors:** Prof. Dr.-Ing. habil. Richard Bamler, TUM
Dipl.-Ing. Nico Adam, DLR

**Cooperation:** DLR Deutsches Zentrum für Luft- und Raumfahrt e.V. in der Helmholtz-Gemeinschaft

# Declaration

This thesis is a presentation of my original work. Wherever contribution of others are involved, every effort is made to indicate this clearly, with due reference to the literature and acknowledgement of collaborative research and discussions.

Munich,
October 15, 2013

Homa Ansari

# Abstract

Coherent Synthetic Aperture Radar (SAR) imagery and its potential in interferometric analysis widened the horizon of earth observation and geodetic measurements in the past decades. Overcoming the error sources in conventional Interferometric SAR (InSAR) techniques, the advanced methods of Persistent Scatterer Interferometry (PSI) and differential SAR Tomography (TomoSAR) have been able to retrieve the geophysical signals of interest from SAR imagery. The key concept in such techniques is constraining the analysis to long time-coherent scatterers available in the imaged scene. Having the main role in such techniques, the coherence of the scatterers must be assessed prior to the interferometric analysis. In this regard, the amplitude time series can be exploited not only to assess the long time-coherency of the scatterers but also to investigate partial stability in a limited time interval of the acquisition period, the latter resulting in detection of temporal coherent scatterers. Exploitation of such temporal scatterers in advanced InSAR methods enhances the information content of the resulting products.

The focus of this thesis is on the investigation of amplitude time series of interferometric stacks in order to detect long as well as partial time-coherent scatterers as the target points for the advanced interferometric methods.

In the framework of this thesis, a generic Bayesian inference software is developed in order to deal with the detection, estimation and model selection involved in the time series analysis. The inference is defined based on the Bayesian network which summarizes the probabilistic dependencies as well as stochastic characteristics of the random variables of a defined model. The formation of such a probabilistic network together with exploitation of numerical approaches in the inversion renders the inference framework highly capable in dealing with any arbitrary inverse problem. The developed software is thus highly adaptable to any InSAR related problem and is capable of handling the non-linearity of the corresponding models.

Using the developed Bayesian inference software, the time-coherence is investigated via estimation of the Signal to Clutter Ratio (SCR) of the scatterers in each resolution cell. The relation between the estimated SCR and phase coherence is assessed through simulations and application to TerraSAR-X data. Based on this result an algorithm is proposed for detection of the temporal coherent scatterers and estimation of the time interval in which the coherence occurs. The proposed algorithm is also evaluated via simulations as well as application to TerraSAR-X data.

**Keywords:** Bayesian inference, Bayesian network, Model selection, Non-linear inversion, Temporal Persistent Scatterer, SCR Estimation, Coherence estimation, Interferometric SAR, Persistent Scatterer Interferometry, TerraSAR-X

# Acknowledgments

The help and support of many individuals has paved the way in my studies and research at DLR-IMF, TUM-LMF and ESPACE, and I wish to acknowledge them here:

First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Richard Bamler not only for providing me with the opportunity to work at his institute but also for giving me the initial motivation and inspiration to choose SAR remote sensing as my specialization.

Secondly, I wish to express my warm gratitude to Nico Adam for his great supervision and endless patience. He taught me the right approach toward problem solving, was always open for questions and discussions and supported me with his invaluable experience and profound knowledge. Working with Nico and his team taught me a lot and was a great honor. I also feel indebted to many of the colleagues of the InSAR team, namely Dr. Ramon Brcic, Alessandro Parizzi, Kanika Goel and Fernando Rodriguez both for the fruitful discussions and the technical support.

I also enjoyed collaboration with the Remote Sensing Technology chair of TUM during the course of my studies and am specifically thankful to Dr. Xiaoxiang Zhu, Dr. Stefan Gernhardt, Dr. Stefan Auer and Yuanyuan Wang for helping me in the introductory steps of my work in the field of SAR remote sensing.

I am also very much thankful to the ESPACE board and lecturers who provided me with much knowledge about state of the art techniques in satellite technology and applications. I specifically wish to thank Prof. Urs Hugentobler and Prof. Michael Schmidt for their encouragement and recommendations in the course of my studies, as well as Prof. Michael Eineder for my first introduction to SAR and InSAR techniques.

Last but not least, I am thankful to my family for their emotional support and to my boyfriend and colleague, Sina Montazeri, both for his encouragement along the way and for the discussions about my ideas and thoughts.

# Contents

# List of Figures

# List of Tables

# Acronyms

| | |
|---|---|
| **AIC** | Akaike Information Criterion |
| **APS** | Atmospheric Phase Screen |
| **BIC** | Bayesian Information Criterion |
| **BN** | Bayesian Network |
| **CDF** | Cumulative Distribution Function |
| **DEM** | Digital Elevation Model |
| **DGM** | Directed Graphical Model |
| **DLR** | German Aerospace Center |
| **DS** | Distributed Scatterer |
| **GENESIS** | GENEric System for Interferometric SAR |
| **GNSS** | Global Navigation Satellite System |
| **i.i.d** | independent identically distributed |
| **InSAR** | Interferometric Synthetic Aperture Radar |
| **KDE** | Kernel Density Estimation |
| **LAMBDA** | Least squares AMBiguity Decorrelation Adjustment |
| **MAP** | Maximum A Posteriori |
| **MDL** | Minimum Description Length |
| **MLE** | Maximum Likelihood Estimation |
| **MMSE** | Minimum Mean Square Error |
| **PDF** | Probability Density Function |
| **PS** | Persistent Scatterer |
| **PSI** | Persistent Scatterer Interferometry |
| **SAR** | Synthetic Aperture Radar |
| **SCR** | Signal to Clutter Ratio |
| **SIR** | Sampling Importance Resampling |
| **SLC** | Single Look Complex |
| **TanDEM-X** | TerraSAR-X Add-oN for Digital Elevation Model |
| **TomoSAR** | SAR Tomography |
| **TPS** | Temporal Persistent Scatterer |
| **RV** | Random Variable |

# Chapter 1

# Introduction

Since its advent in 1980s to date, Synthetic Aperture Radar (SAR) has proved to be a powerful active microwave remote sensing technique in earth observation. Recently, the launch of sub-meter resolution SAR missions such as TanDEM-X, TerraSAR-X and COSMO-SkyMed has opened new vistas to applications of the technique in urban as well as rural areas.

Owing to the SAR coherent imaging, Interferometric Synthetic Aperture Radar (InSAR) approaches exploit the difference of the phase of two complex SAR images acquired at different times or with a slightly different geometry [1]. The time series of interferometric phase provide an invaluable source of information for geodesy and geodetic analysis such as topographic mapping, earth surface motion and structural deformation monitoring. However, exploitation of these time series is limited due to geometrical and temporal decorrelation of the scatterers in the imaged scene [2]. Restricting the time series to highly coherent points in the radar scenes provides a solution to this problem. Such points may be seen as permanent GNSS stations spread densely around the globe, wherever SAR acquisitions are available. Advanced InSAR methods such as Persistent Scatterer Interferometry (PSI) [3] and differentail SAR Tomography (TomoSAR) [4, 5] exploit these points and boost the accuracy of such geodetic measurements.

The contribution of different physical and geometrical features of the imaged scene in formation of SAR complex valued data; and consequently in interferometric phase, complicates the inversion problems involved in the advance InSAR methods. The inverse problems are usually highly non-linear, ill-conditioned and specifically in the presence of non-Gaussian noise. A classical approach to such inversions is to consider appropriate linearization and/or regularization methods; examples of which are found in a variety of employed mathematical tools, ranging from regularized least-squares-based methods to maximum ensemble coherence and compressive sensing. An alternative approach to these mathematical tools is defined in the context of modern Bayesian statistics, where the inverse problem is defined as a probabilistic model and the solutions are sought by maximization of the probabilities. Despite its power, the Bayesian methods are usually analytically intractable. However, the modern numerical approaches provide the solution to the intractability.

The development of a generic Bayesian inference tool, which is adaptable to any kind of InSAR related inversion problem, is thus highly beneficial. Such a generic tool can

provide the only solution to some inverse problems while simplifying the mathematical complexity as well as the computational effort of the others.

The current chapter provides an introduction to the thesis. The motivation behind the current work is expanded in the first section. The sought objectives are later summarized in the second section and finally the outline of the thesis is represented in the last section.

## 1.1 Problem statement and Motivation

The current work has been carried out with two distinct objectives. The initial focus is on the implementation of a comprehensive inference software library based on the Bayesian probability theory, which suffices to cope with an arbitrary InSAR inversion problem formulated in the probabilistic sense. On a second level, the aim is to utilize the developed software package in investigation of coherence of the interferometric phase stack based on the corresponding amplitude. To avoid confusion, each of these goals is treated separately in the following two sections.

### *Bayesian Inference Tool Development*

The experimentally observed data of a process are used in an inference in an attempt to describe its physical reality. The physical reality of the process is usually described by a mathematical model, with the describing elements known as parameters. The retrieval of these parameters from the set of measured data is referred to as inverse problem.

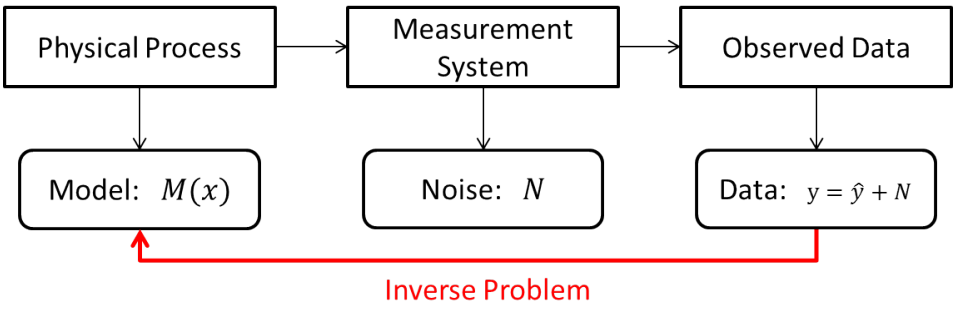



**Figure 1.1: Inverse problem; the physical process is described by a mathematical model $(M(x))$ and retrieval of parameters of the model $(x)$ from the experimentally observed data $(y)$ is sought through the inversion**

Bayesian inference is the most general solution to an inverse problem. Compared to its conventional counterparts, the Bayesian solution is advantageous due to the following capabilities of this approach:

- Integration of prior knowledge about the parameters to the problem set up
- Consideration of non-Gaussian noise in the data
- Handling non-linearity of the forward model without any approximations

- Finding the global optimum solution to the inverse problem without falling into local extremums
- Direct assessment of efficiency of the considered forward model in the light of observed data

In order to develop a generic Bayesian inference tool adaptable to any arbitrary problem, intelligent numerical algorithms are deployed in the implementation. Based on the result of the Bayesian tool, different inference tasks such as detection, estimation and model selection are accomplished.

### *Phase Coherence and Amplitude Time Series Analysis*

The InSAR technique is limited to the coherent conditions where the received signals from the two involved SAR images are correlated [2]. Loss of coherence is known as decorrelation and mainly happens due to the following phenomena: [6]

1. The spectral shift mainly corresponding to spatial separation of the sensors, known as baseline, between the two SAR acquisitions.
2. Temporal change in the position or scattering properties of elementary scatterers inside the SAR resolution cell

In this context, the scatterers in the SAR images are classified into two groups: the Distributed Scatterer (DS) and the Persistent Scatterer (PS); the former is strongly affected by the aforementioned decorrelating signals while the latter is insensitive to such decorrelation and thus is of interest in the advanced InSAR techniques. The temporal phase coherence of the PS can be estimated by:[2]

$$\gamma = \frac{E\{z_1.z_2^*.\exp(-j.\phi_{sys})\}}{\sqrt{(E\{|z_1|^2\}E\{|z_2|^2\}}} \tag{1.1}$$

where $z_i$ denotes the complex value of the interferometric pairs, " $*$ " indicates the complex conjugate operator and $E\{.\}$ is the first moment operator which can be approximated by temporal ensemble average. The coherence determined in this manner considers the interferometric phase variation due to data noise as well as systematic phase deviations ($\phi_{sys}$). The latter is a result of effective range variation which can be attributed to topographic, atmospheric or deformation signals [2].To have an unbiased estimation of the PS temporal coherence, Eq. 1.1 eliminates the systematic decorrelating signals from the interferometric phase; an approach which provides the exact solution but is computationally too expensive to be applicable. Alternatively; on a first level the inexpensive amplitude-based algorithms are applied to infer the Signal to Clutter Ratio (SCR) of the scatterers and estimate the phase coherence by: [6]

$$\gamma = \frac{SCR}{1 + SCR} \tag{1.2}$$

and to detect the PS candidates. These candidates are then analyzed by expensive phase-based time series algorithms in order to retrieve the systematic signals.

In summary, the amplitude time series can be exploited effectively to infer information about phase coherence of the cells and specify the persistent scatterers.

As mentioned, advanced InSAR techniques restrict their interferometric analysis to long time-coherent scatterers in a stack of tens of differential interferograms with one master image. However, a number of points may be found that have partial phase stability in a limited time interval. Such points which are referred to as Temporal Persistent Scatterer (TPS) are so far not integrated into the general framework of advanced InSAR methods such as PSI and differential TomoSAR. Integration of TPSs into these frameworks may increase the density of the points or in extreme cases reveal the changes in the imaged scene which are ignored by these coherent techniques.

The motivation behind this work is to perform time series analysis on the amplitude information of stack of SAR images in order to assess temporal coherence of the radar scatterers. Such analysis gives rise to detection and exploitation of PS and TPS points in the coherent InSAR techniques.

## 1.2 Research Objectives

Based on the stated problems and motivation, the thesis has been carried out with the following objectives:

- **Development of a comprehensive generic software package based on the Bayesian inference:** The software shall be developed in the IDL programming language and must be capable of handling any arbitrary probabilistic problem with as little user interaction as possible. The probabilistic models must be defined in terms of Bayesian networks and the statistical manipulations must be performed based on the defined network.
- **Implementation and assessment of model selection and MAP estimation, based on the developed software**
- **Testing the developed framework with a synthesized inference problem:** A simple linear regression problem is used to test the performance of individual software modules.
- **Implementation and assessment of amplitude time series analysis based on the Bayesian framework:** Model selection is performed on amplitude time series to detect persistent scatterers in the radar scene; signal parameters are estimated for the detected coherent signals and finally the quality of inferred signal parameters in relation to the phase coherence is assessed.
- **Implementation and assessment of Bayesian algorithm for analysis on Temporal Persistent Scatterers:** Partial phase stability of the interferometric stack is detected via amplitude time series analysis, time intervals in which the phase content is stable is specified after the detection.
- **Testing and validation of the developed algorithms through simulated data:** Considering different scattering scenarios, stacks of complex SAR data are simulated to assess the performance of amplitude time series analysis for the two specific cases of PS and TPS.
- **Feasibility study of the algorithms using the real data:** The algorithms are applied to TerraSAR-X data stacks in order to assess the performance of the algorithms in dealing with real life problems.

## 1.3 Thesis Outline

After the introductory chapter, the thesis is continued with two main chapters; dedicated to the Bayesian inference software development and its demonstrative application of amplitude time series analysis, and finalizes with a concluding chapter. The chapters are organized as followed:

Chapter 2 focuses on the development of the Bayesian inference tool. It sets the concepts of the Bayesian inference in the first section while giving the mathematical tools required in handling the numerical approaches toward the inference in the second section. The introduced mathematical and statistical concepts are exploited to form the generic software, the detailed explanation of which is given in the third section. Having the outcome of the developed software as a basis, the model selection and estimation tasks are described based on this outcome to conclude the chapter of Bayesian inference.

Chapter 3 attempts to demonstrate applications of the Bayesian inference software by using amplitude time series for estimation and model selection. The PSI framework is reviewed in the first section in order to clarify the relevance of the amplitude time series analysis to this specific InSAR approach; with the motivation of PS and TPS detection expanded in the continuation of this section. The observed SAR signal statistical characteristics are reviewed in the second section of the chapter, as required background information. The two applications of PS and TPS detection are then expanded, justified and validated with both simulated and real data in the two final sections.

Finally the thesis is summarized in Chapter 4 followed by concluding remarks and the outline for further research works.

# Chapter 2

# Bayesian Inference Software Development

This chapter is dedicated to the Bayesian inference and the developed software for handling the probabilistic models defined based on this approach. The two first sections provide the necessary concepts and definitions; with section 2.1 dedicated to the Bayesian inference and section 2.2 to the mathematical algorithms required in the developed software. The step by step explanation of the software is provided in section 2.3 and the ultimate solution to the inference task is finally discussed in section 2.4. The concepts and methodologies are further explained by a simple linear regression example which is defined and resolved throughout the sections.

## 2.1 Introduction to Bayesian Inference

The key concept in the inference can be summarized in one single word: *probability*. The probability theory is interpreted by two broad views: the classical or frequentist view that interprets the probability as the frequency of random repeatable events as opposed to the Bayesian view in which probabilities are understood as the quantification of uncertainties. This section provides basic concepts of Bayesian probability theory and its solution.

The section starts with the definition of Bayes' theorem in sub-section 2.1.1, in sub-section 2.1.2 the probabilistic models are introduced via the Bayesian networks. Probabilistic manipulations in the Bayesian networks are explained in sub-section 2.1.3 and the last sub-section gives a general overview toward the numerical approaches to such manipulations in the network.

### 2.1.1 Bayes Theorem

In the gist of it, the probability theory can be expressed in terms of two basic rules of sum and product. The first is axiomatic, the belief in the probability of how much a proposition ($X$) is true implicitly specifies the probability of how much it can be wrong, i.e.: [7]

$$\textit{Sum rule}: \qquad P(X|M) + P(\bar{X}|M) = 1 \tag{2.1}$$

The second is more complex: if the belief about one proposition ($Y$) is known and the probability of occurrence of the second proposition ($X$) given that the first one has occurred is also stated, then the probability of the joint occurrence of the two propositions is implicitly stated as well. i.e.: [7]

$$Product\ rule: \qquad P(X,Y|M) = P(X|Y,M) \times P(Y|M) \tag{2.2}$$

In this content $P$ as the probability, $\bar{X}$ denotes the proposition that $X$ is false, the vertical bar "|" is the conditionality symbol which specifies the *conditional probabilities* and the comma is the conjunction between the propositions which indicates the *joint probabilities*. Both propositions are made conditional on $M$, which is the link to the underlying background information of the propositions at hand and will be further referred to as the underlying model.

From the product rule, together with the *symmetry property* $p(X,Y) = p(Y,X)$, the following relation between conditional probabilities is immediately obtained:

$$(X|Y,M) = \frac{p(Y|X,M) \times p(X|M)}{p(Y|M)} \tag{2.3}$$

This relation is the well-known Bayes' theorem. The strength of this theorem is in its ability to turn the conditionality between the propositions around. To clarify this statement, let us assume a simple linear regression problem with two parameters:

$$\bar{Y} = \bar{M}\bar{X} = x_0 + \bar{M}x_1 \tag{2.4}$$

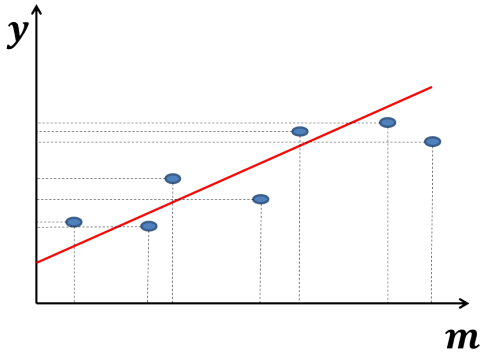



**Figure 2.1: Linear regression with two parameters: offset and slope; the dots are the observed values and determination of the two parameters of the red line is desired**

Fig. 2.1 is a sketch of this defined regression problem. In the defined setup, $X$ is the vector containing the offset ($x_0$) and slope ($x_1$) of the linear model as the sought parameters of the problem, $\bar{Y}$ and $\bar{M}$ are vectors containing the observed data and the sampling points on which the data is observed, respectively. $\bar{M}$ is known with certainty thus is referred to as the model structure. In this set up the defined theorem is expressed by the following proportionality:

$$p(parameters|data,model) \propto p(data|parameters,model) \times p(parameters|model) \tag{2.5}$$

The power of the Bayes' theorem lies in the fact that it relates the quantity of interest, the probability that the parameters of the linear model are correct given the data, to the term which has a better chance to be assigned i.e. the probability under which the data is measured under the certain linear model $M$ with the given parameters $X$.

Each term of the Bayes' theorem is known with a specific name. $p(parameters|model)$ is called the prior. It represents the knowledge about the truth of the parameters prior to observation of data. This prior knowledge is further modified by the experimental data through the so-called likelihood probability or $p(data|parameters, model)$ and yields the posterior probability $p(parameters|data, model)$. Posterior probability determines the uncertainty about the parameters of the assumed model in the light of the data [7]. In Eq. 2.5 the theorem was explained with the proportionality, the equation is complete by considering the normalization constant $p(data|model)$, known as the evidence. The evidence is irrelevant for some inference tasks such as parameter estimation while it may have a key role in others such as model selection. sub-section 2.3.3 and sub-section 2.4.2 focus on the concept and the importance of evidence.

### 2.1.2 Bayesian Networks

As discussed before, the Bayes' theorem encapsulates the uncertainties about the different elements of a mathematical model thus leading to a probabilistic model to handle the solution of its' mathematical counterpart. To cope with the probabilistic models it is highly advantageous to summarize them in terms of diagrammatic representations of probability distributions known as probabilistic graphical models. The main advantages of such diagrams are:[8]

- Simple visualization of the structure of the complex probabilistic model
- Provision of information about the properties of the model, such as conditional dependencies of the variables, through inspection of the graph
- Expression of the complex computations required to perform inference in terms of graphical manipulations

The graphical models comprise two basic elements: nodes, as a representative of Random Variable (RV), and links, as a connection between the RVs indicating the dependencies. Bayesian Network (BN), also known as Directed Graphical Model (DGM), is special type of the probabilistic graphical models in which the links of the graphs have particular directionality.[8] To start with the Bayesian networks, consider the following three RVs and the respective defined joint probability among them:[8]

$$p(a, b, c) = p(c|a, b)p(b|a)p(a) \tag{2.6}$$

As it is evident, the joint probability is decomposed into the conditional probabilities through the product rule thus forming a simple probabilistic model. This decomposition is captured by the simple Bayesian network in Fig. 2.2. Each conditional probability of the left hand side of Eq. 2.6 is captured by directed link between the RVs involved. For

instance, the conditional probability $p(b|a)$ adds the link directed from the parent node a to the child node b and so for the $p(c|a,b)$, where the two links from the two parents a and b to the child c exhibit the probabilistic dependency of this node on its parents.

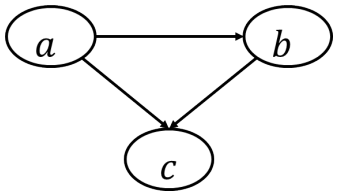



**Figure 2.2: A simple Bayesian network; the joint probability among the three random variables is defined by the conditional probabilities represented by directional links in the graph [adopted from Bishop, 2006]**

To fully describe the probabilistic model, each node is allocated with a Probability Density Function (PDF) indicating the uncertainty or belief associated with the RV. The PDF $f(x)$ is defined as a function which describes the relative probability by which a random variable $(x)$ takes on a given value $(x_0)$:[9]

$$P(x_0) = \int_{-\infty}^{x_0} f(x)\, dx. \tag{2.7}$$

Let us now return to the linear regression example of Eq. 2.4 and form the Bayesian network of the corresponding probability model. Starting with the mathematical system of equations:

$$y_i = x_0 + m_i x_1; \qquad i = 0 \ldots N \tag{2.8}$$

The corresponding probabilistic model for the joint posterior of the parameters $x_0$ and $x_1$ in the light of a single data point is followed by:

$$p(x_0, x_1|y_i, m_i) = \frac{p(y_i|m_i, x_0, x_1) \times p(x_0, x_1|m_i)}{p(y_i|m_i)} \tag{2.9}$$

Each single data point includes new information to the model. The new information can be interpreted as additional conditional probability which helps the joint posterior probability to be more concentrated in its truth. Applying the product rule to cope with the conditionality added by N data points, the joint posterior probability yields:

$$p(x_0, x_1|\bar{Y}, \bar{M}) = \prod_{i=0}^{N} p(x_0, x_1|y_i, m_i) = \frac{\prod_{i=0}^{N} p(y_i|m_i, x_0, x_1) \times \prod_{j=0}^{1} p(x_j|\bar{M})}{\prod_{i=0}^{N} p(y_i|m_i)} \tag{2.10}$$

The BN corresponding to the probability model defined by Eq. 2.10 is given in Fig. 2.3. The mathematical model in Eq. 2.8 justifies the direction of the links in the network: each data point $(y_i)$ is observed at a grid point $(m_i)$ and is related to the offset and slope of the linear model $(x_0, x_1)$. To complete the network, appropriate probability functions must be assigned to each node. There are three types of PDF defined for the current network:

- Uniform: assigned to the parameter nodes as the prior PDF. This non-informative PDF describes a constant probability:

$$f_U(x) = U(a,b) = \begin{cases} 0, & x < a \\ \frac{1}{b-a}, & a < x < b \\ 0, & x > b \end{cases} \tag{2.11}$$

  The limits of this PDF shall be set to the expected range for the corresponding RV.

- Gaussian: assigned to the data nodes as the likelihood PDF. The Gaussian is defined as:

$$f_G(x) = N(\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \; e^{(\frac{-(x-\mu)^2}{2\sigma^2})} \tag{2.12}$$

  The two parameters of the PDF $(\mu, \sigma)$ are set to the observed value of data $(y_i)$ and the noise of the measurement process, respectively.

- Dirac: assigned to the model structure nodes to show the certainty in the value given to the node. The Dirac PDF is defined as:

$$f_D(x) = U(a,b) = \begin{cases} 1, & x = x_0 \\ 0, & x \neq x_0 \end{cases} \tag{2.13}$$

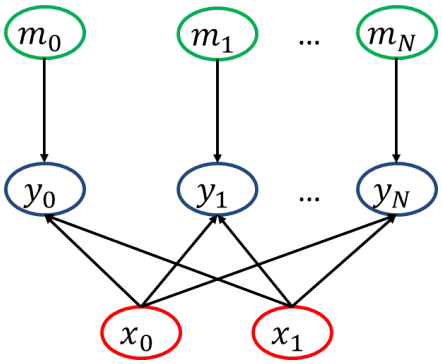



**Figure 2.3: Bayesian network of linear regression with two parameters; the green, blue and red nodes represent the model structure, the observed data and the desired line parameters, respectively**

By inclusion of the probability functions to the nodes the Bayesian network is complete.

### 2.1.3 Inference on Bayesian Network

The goal of the inference is to calculate the posterior probabilities for any nodes of interest in the network. This calculation is possible by propagation of the local probabilities through the network from the parent nodes down to the children. As a clarification of the probability propagation, consider joint probability defined between two random variables $x$ and $y$:

$$p(x,y) = p(y|x)p(x)$$

Suppose that the y nodes are observed and the x node is the target of the inference for which the posterior probability is desired. The likelihood is assigned to y and the prior to the $x$ nodes. The posterior probability is sought through the Bayes' theorem and in the following three steps (Fig. 2.4):

1. Propagation of the probability from parent to child, equivalent to evaluation of the product rule of probability, achieved by multiplication of prior and likelihood i.e. $p(x) \times p(y|x)$
2. Marginalization of the probabilities in the child node, equivalent to evaluation of the sum rule of probability, defined by summation of the joint probability over all random variables except for the variable of interest, i.e. $p(y) = \sum_x p(y|x)p(x)$
3. Reverse propagation from the child to the parent node, the joint probability is now expressed in terms of $p(y)$ and $p(x|y)$, i.e. $p(x|y) = \frac{p(y|x)p(x)}{p(y)}$

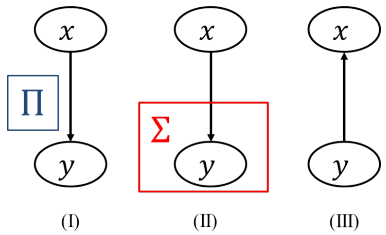



**Figure 2.4: Graphical probability propagation in Bayes' theorem; (I) local probability propagation form the parent to the child as the nominator of the Bayes theorem (II) marginalization of the probabilities in child node as the denominator of the Bayes theorem (III) reverse propagation from child to parent to get the posterior probability**

Inference on any arbitrary Bayesian network is described in the same propagation manner and by the same simple product and sum rules regardless of the complexity and the size of the network. The only difference will be the number and levels of parent nodes for the child which impose more components to the products but makes no changes in the basic manipulations.

### 2.1.4 Numerical Approaches toward Bayesian Inference

Dependent on the complexity of the probabilistic model and the probability functions involved, it is possible to find analytic closed form solution to the inference. It might also be possible to find deterministic approximations to the posterior probabilities of interest [8, 10], but in its most generic form the Bayesian inference is an intractable problem. To set up a generic inference framework which is adoptable to any desired probabilistic model, best is to consider the approximation methods based on numerical sampling also known as Monte Carlo technique. The strength of such generic approach is its flexibility toward intractable as well as the well-behaved models. While providing the only solution to the former, Monte Carlo technique also simplifies inference on the latter by omitting complicated mathematical approximations.

Monte Carlo methods are classically defined as any method which solves a problem by generating suitable random numbers and observing that fraction of the numbers obeying some property or properties [11]. In the context of probabilistic models, the random numbers must be generated from a desired probability distribution. A wide variety of the Monte Carlo sampling methods for investigation of PDF is available in the literature; ranging from the most basic methods such as rejection sampling to the most advanced ones such as Markov Chain Monte Carlo [12, 13, 14]. The difference between such methods lies on the level of intelligence by which the informative part of the PDF is sampled. The trade-off for such smart algorithms is their complexity, expensive computational burden and slow performance.

Among the sampling algorithms the so called Sampling Importance Resampling (SIR) method is favored for the developed software, due to its near on-line performance and efficiency in concentration on the informative section of the PDF [15]. The detailed explanation of this method is in the scope of sub-section 2.2.3.

The role of the sampling methods in the Bayesian networks is initialization of the probability propagation task. The sampling occurs at the parent nodes where single values are drawn from the PDF assigned to the node ($x_0 \sim f(x)$). Having the sampled value, its probability is derived by evaluation of node's PDF ($p = f(x_0)$), the sampled value weighted by its probability is referred to as *particle*. The particles of the parental nodes move along the links of the graph, helping in evaluation of likelihoods and propagation of the probabilities in the network. Fig. 2.5 is an illustration of sampling method in the Bayesian network defined for the linear regression example (Fig. 2.3). In this figure the red box, blue and green curves respectively represent the uniform prior PDF of the parameter space, the Gaussian likelihood PDF of a single data node and finally the resulted joint posterior PDF of the parameters in the light of the single data. The samples, shown by the red dot, are drawn from the uniform PDFs of the parameter nodes; moving to the data node, the sample evaluates the likelihood, shown by the blue dot and finally the posterior probability is calculated as the product of prior and likelihood probabilities.

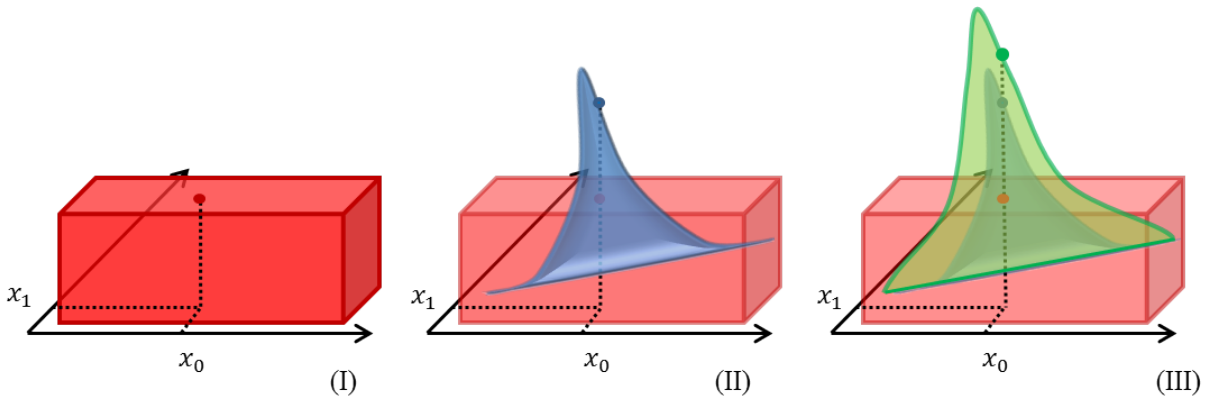



**Figure 2.5: Monte Carlo sampling in evaluation of posterior probability for the linear regression example; (I) sampling from the parameter space, (II) evaluation of the likelihood of a single data for the sampled parameters, (III) calculation of the joint posterior probability for the sampled parameters under the light of the single data**

Repeating the sampling scheme, over the prior parameter space and for all the observed data, leads to creation of the posterior probability.

## 2.2 Mathematical and Statistical Background

The applied mathematical and statistical algorithms in the software are explained in this section to provide a comprehensive understanding of the software details.

### 2.2.1 Monte Carlo Integration

Monte Carlo integration is a numerical method for integration of a complicated function over a specific defined domain and is mostly beneficial in solution of multi-dimensional integrals. Suppose that the integration of M-dimensional function $f(x)$ on a defined domain $V$ is desired, the domain is $M$ dimensional, and the function is analytically easy to evaluate but complicated to be integrated:

$$I = \int_V f(x)\, dx.$$

The solution is to draw random vectors $(x)$ from the defined domain and evaluate the function for the sampled vector. Then an estimate of the integral is given by:

$$\hat{I} = \bar{f}V = \frac{V}{N}\sum_{i=1}^{N} f(x_i) \tag{2.14}$$

with $V$ as the M-dimensional integral of the domain from which samples where generated and $N$ as the number of samples considered in the estimation.

### 2.2.2 Inverse Transform Sampling

Inverse transform sampling allows the generation of random numbers from a desired distribution. To describe this method it is necessary to introduce the Cumulative Distribution Function (CDF) of a PDF. The CDF is defined as:

$$F(x) = \int_{-\infty}^{x} f(x')\, dx'. \tag{2.15}$$

Thus the following equation relates the PDF and CDF:

$$f(x) = \frac{dF(x)}{dx} \tag{2.16}$$

CDF is proved to be monotonously increasing with its range limited to the unit interval:[9]

$$\lim_{x\to-\infty} F(x) = 0 \quad ; \quad \lim_{x\to+\infty} F(x) = 1 \tag{2.17}$$

It is also apparent that:

$$\text{If } \; z = F(x) \; \text{ then } \; x = F^{-1}(z) \tag{2.18}$$

Thus having generated a random number in the unit interval ($z$), one is able to calculate the corresponding random value $x$ only through a simple inversion of the CDF. The resulted $x$ follows the desired distribution $f(x)$. There are two necessities to the mentioned method:

1. Generation of a random number in the unit interval
2. Solution of the CDF integral and its inverse

The first is not a limit to the method. The pseudo random number generators, although subtle in concept, are available in most of the software environments. The second one however limits the utilization of this algorithm in cases where the CDF is not analytically tractable or invertible. In such cases the alternative is the Monte Carlo based method, which is introduced in the following sub-section.

### 2.2.3 Sampling Importance Resampling

The SIR is a Monte Carlo based approach for random number generation from a desired probability distribution function $f(x)$ for which the inverse transform sampling methods are not applicable [8]. In this case a close distribution function to $f(x)$ is considered ($g(x)$) from which the direct sampling is possible. Having chosen this so-called proposal distribution, there remain two stages to the sampling scheme. In the first stage, $N$ samples are generated form the proposal distribution and are later weighted by the normalized ratio of the two probabilities:

$$z_i \sim g(x); \quad w_i = \frac{f(x_i)/g(x_i)}{\sum_N f(x_i)/g(x_i)} \tag{2.19}$$

As for the next stage, a second set of $N$ samples is drawn from the discrete distribution $[z_1, z_2, \ldots, z_N]$ with probabilities given by the weights $[w_1, w_2, \ldots, w_N]$. The resulting $N$ samples are asymptotically drawn from the desired distribution.

To be applied to imitation of posterior PDF in Bayesian network, the described SIR method is adopted in the following way:

1. The proposal distribution gives its' place to the prior PDF of the parental node
2. The importance weights are substituted by the propagated probabilities around the network

The details will be provided later in the description of software packages in sub-section 2.3.2 and sub-section 2.3.3.

### 2.2.4 Systematic Resampling

The resampling involves generating a set of samples from an approximate discrete representation of the desired PDF represented by particles:

$$f(x_i) \approx \sum_{i=1}^{N} w_i \delta(x - x_i) \tag{2.20}$$

In here, $w_i$ represents the weight of the particle while $x_i$ indicates its value. The basic idea of resampling is to eliminate the particles which have small probabilities and generate particles with higher probability, leading to concentration on the massive part of the PDF represented by discrete particles. The resulting resampled set will be independent identically distributed (i.i.d) samples from the discrete density of Eq. 2.20 with the corresponding weights of $w_i = \frac{1}{N}$ [15].

There are a number of resampling schemes available in the literature, from which the systematic resampling outperforms the others [16]. Systematic resampling is based on the previously discussed inverse transformation theory Eq. 2.18. The initial $z$ values are chosen systematically by:

$$z_k = \begin{cases} \frac{(k-1)+u}{N}, & 1 \leq k < N \\ 1, & k = N \end{cases} \tag{2.21}$$

where $u$ is a single random number, generated from uniform distribution in the unit interval, $k$ corresponds to the index of the resampled particle and $N$ represents the sample size.

The CDF is derived applying the Monte Carlo integration on the discrete density function $(f(x))$:

$$F(x) = \frac{N^{-1}}{x - min(\bar{x})} \times \sum_{x'=min(\bar{x})}^{x} f(x') \tag{2.22}$$

with $\bar{x}$ as the vector containing the initial samples. Finally the resampled values are derived from the inversion transformation: $x'_k = F^{-1}(z_k)$

### 2.2.5 Optimum Number of Samples

The number of samples in the Monte Carlo techniques has a twofold effect on their performance. Increasing the number of samples improves the accuracy of the imitation process on the one hand while imposing more computational burden to the algorithm on the other hand. To set a balance, the number of samples is related to the desired precision for the approximation. The precision is defined as the deviation of the approximated from the exact analytical mean of the target PDF. The analytical mean is obtainable from: [17]

$$E_p(g(x)) = \int_{-\infty}^{+\infty} g(x) f(x)\, dx. \tag{2.23}$$

and is approximated with the generated samples from the PDF by:

$$\hat{E}(g(x)) = \frac{1}{N} \sum_{i=1}^{N} g(x_i) \tag{2.24}$$

Finally the precision of the approximation is defined as:

$$\epsilon = \hat{E}(g(x)) - E_p(g(x)) \tag{2.25}$$

The precision may be limited by different criteria such as Bernoulli, Chernoff and Chebyshev's bounds [18]. The Chebyshev's bound is favored among the others since it is affected by the type of the distribution function and defined as: [17]

$$q(N, \epsilon) = \frac{var(g(x))}{N\epsilon^2} \tag{2.26}$$

Assuring that $q(N, \epsilon) < \delta$, the optimum number of samples for approximation of the target PDF is derived as:

$$N \geq \frac{\sigma^2}{\delta\epsilon^2}; \quad \sigma = \sqrt{var(g(x))} \tag{2.27}$$

In the right hand side of this inequality, the numerator is the empirical second moment of the target PDF which is also known as the variance of the sample ($\sigma^2$); later it is seen how this value is inferred from the learning step of the developed algorithm. $\delta$ and $\epsilon$ denote the confidence level and the precision of the approximation, respectively. In summary, the inequality states that: Given $N$ number of samples, the empirical PDF resulted from sampling is not more than $\epsilon$ percent away from the true PDF. This statement is assured by confidence level of $1 - \delta$ percent.

### 2.2.6 Estimation of Probability Density Function

Throughout the chapter, two representations of the PDFs have been introduced:

1. Densities having specific functional form governed by a number of parameters, which are called parametric densities e.g. Eq. 2.12

2. Densities reconstructed by particles, which are referred to as particle densities and were formulated by Eq. 2.20

The former provides a continuous function while the latter is a discrete form comprised of irregularly sampled values carrying probabilistic weights. Although initially useful for investigation of complicated PDFs, the particle representation is not appropriate for further statistical analysis. Thus an intermediate regularly sampled representation is provided as a link between the particle and parametric forms by conversion of the particles. As will be discussed at the end of this chapter, this conversion is also beneficial in ultimate parameter estimation of BN.

The goal here is to estimate the probability density function ($f(x)$) given a set of its observed data ($x$). Having the discrete PDF represented by particles, the first necessity is to generate data from the discrete PDF to provide the input data ($x$) for this estimation. As discussed in sub-section 2.2.4, systematic resampling allows us to draw a set of samples from arbitrary discrete PDFs. Fig. 2.6 illustrates scheme of the described concept. Note that the massive part of the density function results in higher sampling frequency of the data points. This is in fact this frequency of the data points which enables the estimation of the target PDF. The first possibility in estimation of PDF is the histogram method.

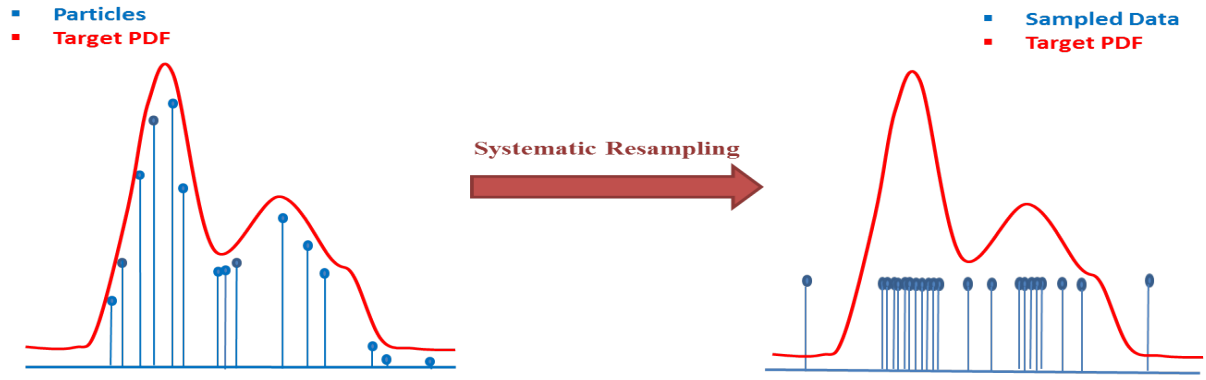



**Figure 2.6: Systematic resampling; conversion from the particle representation of PDF to the Density representation where the density of the samples indicates the probability function**

Standard histograms simply partition the given data ($x$) into distinct bins of width $\Delta$ and record the frequency of the portion of data falling to each bin [8]. Although straight forward in practice, the histogram method is inadequate in application. One obvious problem is that the estimated density has discontinuities that are due to the bin edges rather than any property of the underlying distribution that generated the data. More appropriate method is estimation via moving smoothing kernels. This method which is known as Kernel Density Estimation (KDE) can be formulated as: [8]

$$\hat{f}(x) = \frac{1}{Nh}\sum_{i=1}^{N} k(\frac{\hat{x}_i - x}{h}) \tag{2.28}$$

In this formulation, $k(.)$ is the moving kernel, $h$ the smoothing bandwidth and $\hat{x}_i$ the regularly spaced points defined in the range spanned by the PDF. There are a number of kernels suggested for KDE; such as Gaussian, Uniform, Bi-weight, Epanechnikov etc., among which the last one have proved to be optimum [19]. Epanechnikov kernel is defined as:

$$k = \begin{cases} \frac{3}{4}(1 - (\frac{\hat{x}_i - x}{h})^2), & |\hat{x}_i - x| \leq h \\ 0, & |\hat{x}_i - x| > h \end{cases} \tag{2.29}$$

Fig. 2.7 depicts the shape of the defined kernel as well as schematic representation of the KDE process.

The choice of the kernel and its bandwidth is critical. The bandwidth of the kernel acts as a smoothing parameter. In its extreme case the wide bandwidth leads to loss of information while short bandwidth leads to discontinuities in the resulted PDF. Setting a balance is possible by seeking the optimum bandwidth which preserves the shape while smoothing the spikes of data point. This bandwidth is defined through minimization of the *Asymptotically Mean Squared Error (AMSE)* between the estimated and the exact PDF, the resulted optimum bandwidth is thus derived as: [19]

$$h_{opt} = 2.34.\hat{\sigma}.N^{-\frac{1}{5}} \tag{2.30}$$

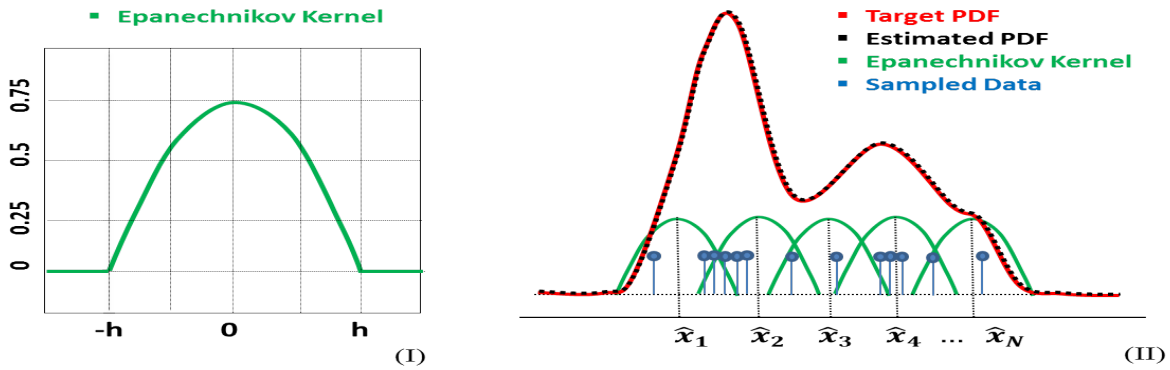



**Figure 2.7: Kernel density estimation; (I) optimum kernel for KDE process, (II) schematic representation of KDE as the moving smoothing kernel on the data spikes**

in which $\hat{\sigma}$ is an estimate of the standard deviation of the data set and $N$ is the number of data points. A range of assumptions and simplifications have been considered in derivation of this optimum value which may not necessarily be held in any arbitrary cases. Still in such cases the suggested bandwidth is close to optimum and a safe choice to apply. [19]

## 2.3 The Developed Software Package

Having introduced the background concepts and mathematical tools, the aim in this section is to bring about a comprehensive framework for construction and solution of the probabilistic models. The software is divided into four packages:

1. Creation of the Bayesian Network
2. Learning of the network
3. Importance Resampling
4. Posterior PDF Estimation

Each of these packages is explained in one of the following sub-sections. The defined linear regression problem of section 2.1 is revisited in each section as a demonstrative example.

### 2.3.1 Creation of the Bayesian Network

The first package is responsible for forming the structure of the Bayesian network from the user defined entities. The network is comprised of set of nodes categorized in two basic types:

- Independent Random Variables: known as parent nodes of graph, the existence of which is independent of any higher level nodes; e.g. the parameters and model structure in the linear regression model
- Dependent Random Variables: the existence and probability of which is conditional on a set of parent nodes; e.g. the observed nodes in the linear regression model

The independent nodes are further divided into two subcategories: The "Known" nodes discriminated by level 1 and the "Latent variables" indicated by level 2 in the software package. The value of the former is known with certainty while the later must be determined with sampling from its respective PDF.

Additional information about the links of the graph and the probabilities assigned to the nodes are also integrated into the node structure. Fig. 2.8 depicts the structure of the resulted network from the current module.

As it is seen the network is comprised of set of random variable defined by classes. Each class contains the attributes: node name and type, PDF information, dependency information, with the latter only relevant to the dependent nodes.

As will be seen later in this section, such structure allows node-specific manipulations which are only relevant to defined levels of node.

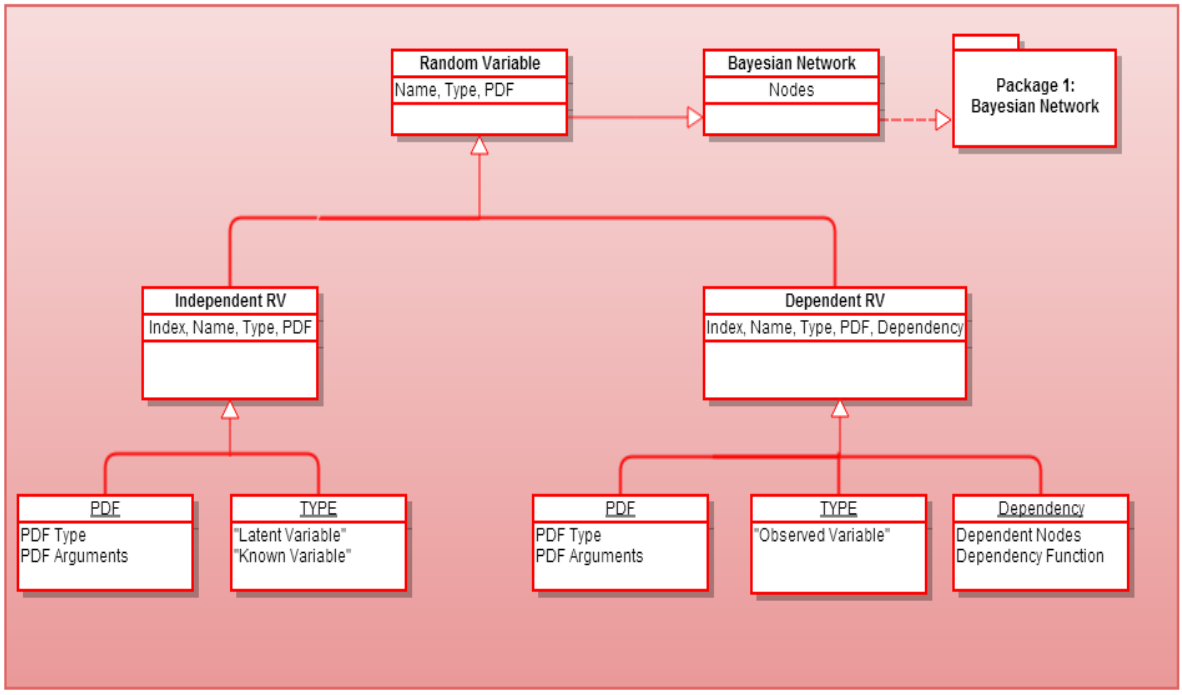



**Figure 2.8: Software package 1: Bayesian network; organization of the user defined probabilistic model by directed graphical model**

### 2.3.2 Learning

The first step toward the inference is to investigate the probabilistic model by particles. This phase can be regarded as the "Sampling" part of the SIR algorithm where the relevant random samples and their probabilistic weights are generated.

The learning is initialized by a random walk through the network and drawing random samples from the independent RVs. In the second stage, the random samples help in evaluation of the likelihood PDF of the dependent nodes. Having the random samples together with the prior and likelihood probabilities, the product rule is conducted to give the a posteriori likelihood of the independent samples. So far the particles together with their non-normalized weights are resulted. From the evidence determination module, the normalizing constant is calculated in the next step and the a posteriori weights are normalized. The result will be the joint posterior PDF over the parameter space.

The joint posterior PDF is the input of the next level where the attempt is on retrieving two specific types of information about the latent RVs:

- The optimum number of random samples for posterior reconstruction
- The range of the target parameters which lie in the informative part of the posterior

The optimum number of samples is obtained from Eq. 2.27, with $\sigma^2$ defined as the empirical variance of the particle values and $\epsilon$ being the desired precision of the estimation.

Retrieval of information about the boundaries, where the informative part of posterior PDF is located, is achieved in the following steps:

- Marginalization of the joint posterior PDF with respect to the parameter of interest
- Calculation of the CDF from the marginalized posterior
- Resampling marginalized PDF given its CDF
- Taking the first and last element of the resampled set as the informative area of the PDF

The outcome of the package would then be the determined number of particles as well as the boundary to the informative part of latent variables; Fig. 2.9 provides a diagram to give a more comprehensive overview of the package.

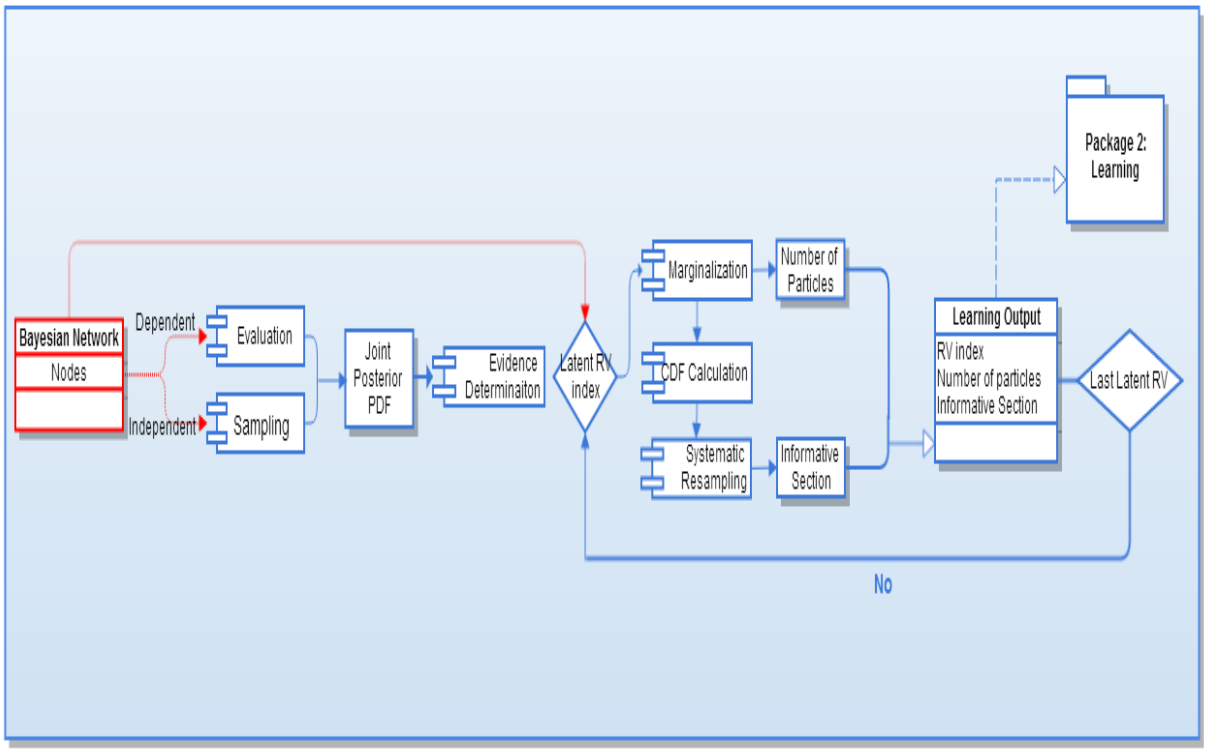



**Figure 2.9: Software package 2: Learning; initial sampling and propagation of the probabilities in the network to derive the initial information about the target parameters**

The linear regression problem is taken here as an example to demonstrate the package. Having drawn the particles in the network and propagated the probabilities from prior to posterior level, the discrete joint posterior density ($f(\bar{X}|\bar{Y},\bar{M})$) is obtained over the parameter space of the model ($\bar{X} = [x_0, x_1]$); Fig. 2.10 shows the resulted joint PDF.The particle joint posterior is marginalized with respect to each of the single parameters. The marginalization is formulated as:

$$f_{x_i} = f(x_i|\bar{Y},\bar{M}) = \int_{\Omega\in[x_j|j=0,\dots,K;j\neq i]} f(\bar{\Omega}|\bar{Y},\bar{M})\, d\Omega \tag{2.31}$$

In here marginalized PDF of each single parameter is denoted by $f_{x_i}$ and $\Omega$ is the parameter space in the absence of the parameter with respect to which the PDF is marginalized.

As discussed before, this integration is numerically evaluated by Monte Carlo integration, i.e.:

$$f_{x_i} = f(x_i|\bar{Y}, \bar{M}) = \prod_{j=0,\ldots,K;j\neq i} \frac{1}{max(x_j) - min(x_j)} \times f(\bar{X}|\bar{Y}, \bar{M}) \tag{2.32}$$

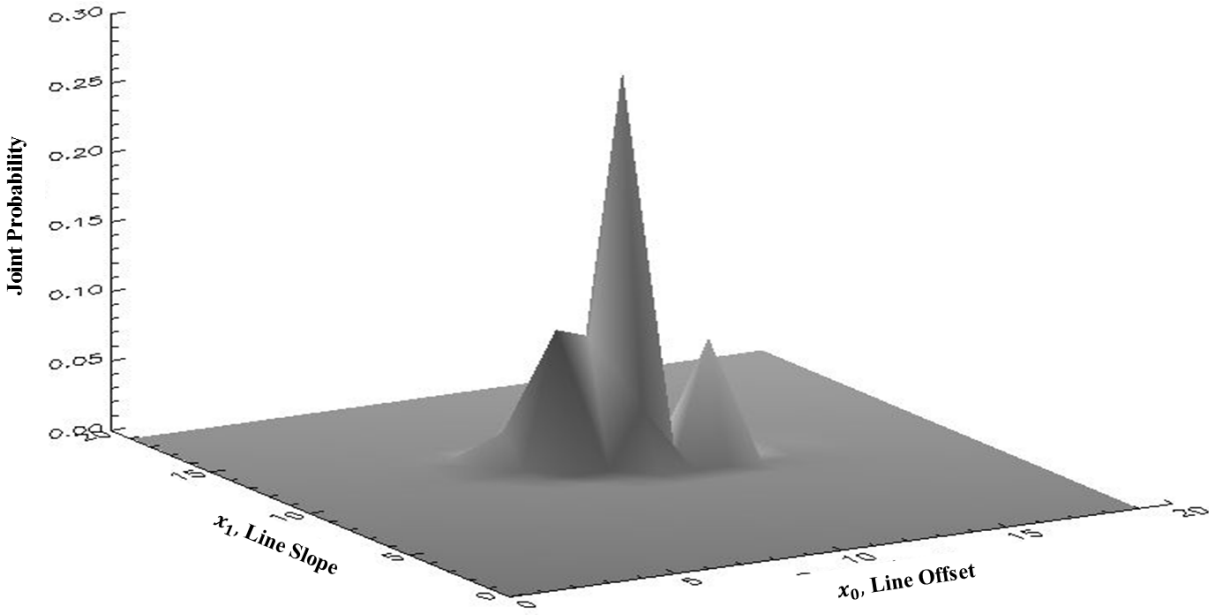



**Figure 2.10: Joint posterior probability of the linear regression example; initial result from the learning step**

The resulted marginalized PDF is then integrated using Eq. 2.22 to give the cumulative distribution function. Fig. 2.11 shows the resulted marginalized density function and its corresponding CDF of the slope parameter ($x_1$). As it is seen the prior range is defined as $[0, 20]$, while the relevant part of this prior spans a small portion of this range. This unsuitable prior is the reason for steepness of the CDF.

This problem can be solved by systematic resampling. Recall form the sub-section 2.2.4 that the resampling will eliminate particles with negligible probabilities. Therefore the maximum and minimum of the resampled set imply the borders to the "massive" area of the discrete PDF. If the learning process is repeated with this informative area as the prior, then the imitated PDF will be more informative. Fig. 2.12 represents the outcome of PDF concentration resulted from applying the described strategy.

Given that the number of particles was the same from Fig. 2.11 to 2.12, it is seen that the applied strategy results in creation of dense particles in the massive part of the PDF via limiting the prior range for the particles.

The repetition of the learning step with the informative part of prior can be regarded as the "Importance Resampling" step of the SIR algorithm and gives an appropriately sampled particle PDF of the joint posterior PDF to be investigated further in the inference procedure.

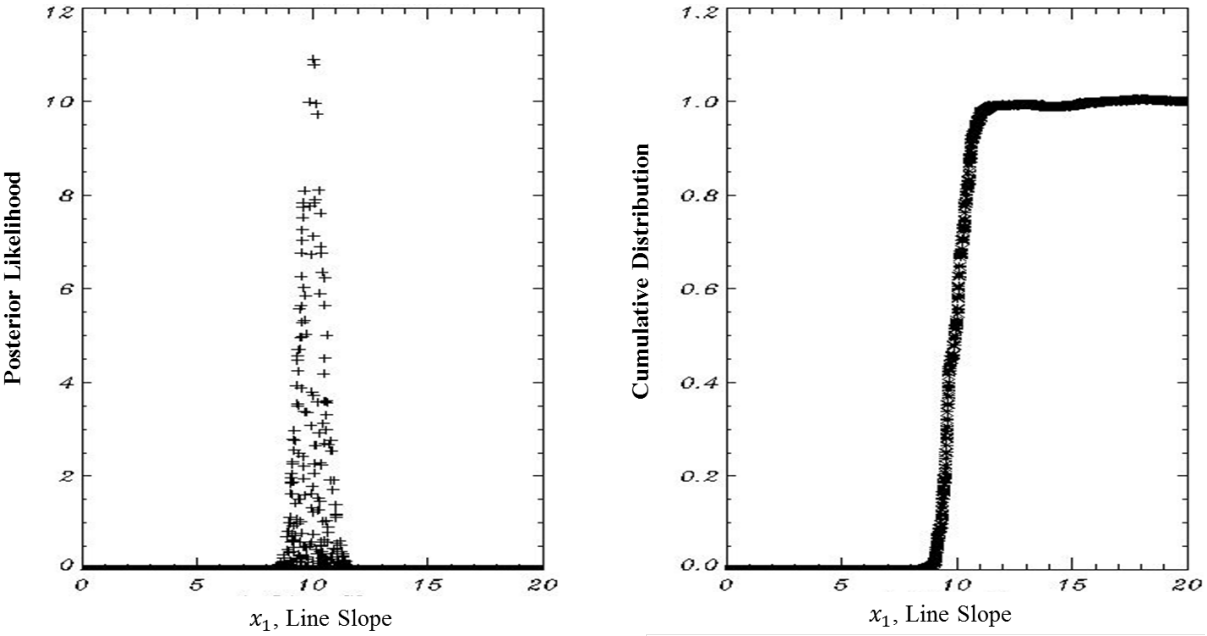



**Figure 2.11: Likelihood and cumulative distribution function of the line slope in linear regression example; initial result from the learning step**

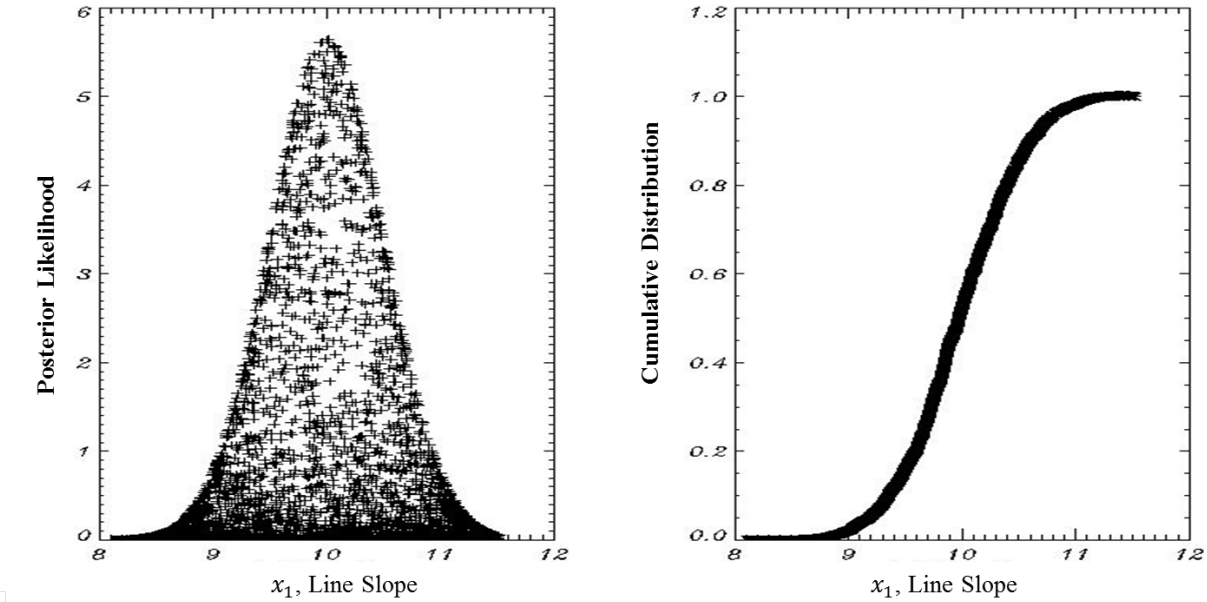



**Figure 2.12: Likelihood and cumulative distribution function of the line slope in linear regression example resulted from limitation of the prior range; using the same number of particles the imitated PDF is more concentrated compared to the initial results**

### 2.3.3 Importance Resampling

The purpose in the current package is to finalize the numerical investigations of the posterior and prepare the inputs to the inference task. The Bayesian network is once again investigated in this package, this time by consideration of the optimum number of particles which are determined for each of the latent variables and by limiting the sampling to the informative section of the posterior. Fig. 2.13 describes the package. Due to the

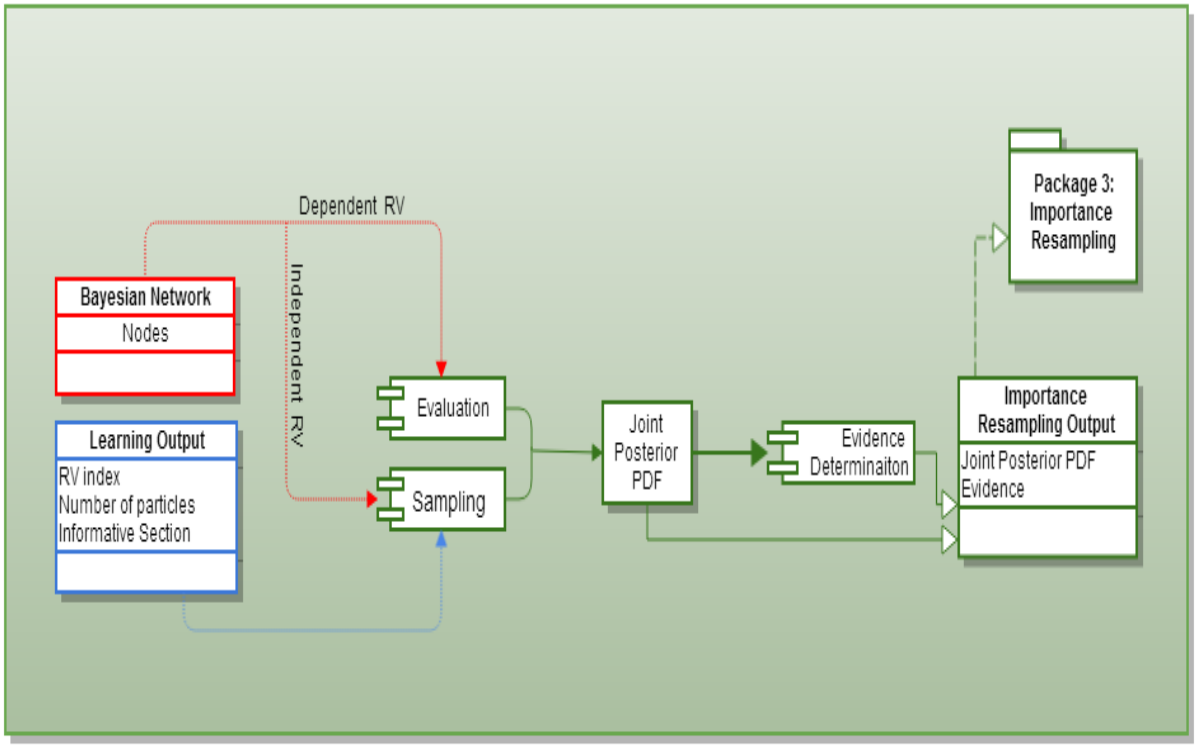



**Figure 2.13: Software package 3: Importance resampling; repetition of sampling using the optimum number of particles and the informative priors to improve the sampling results**

importance of the evidence, as the merit index of the model and the normalizing constant of the posterior PDF, the rest of this subsection is dedicated to the explanation of the evidence determination module.

In its explicit form, evidence is defined as the solution of the following multidimensional integral over the latent random variable space of the network:

$$E = f(\bar{Y}|\bar{M}) = \int_{\Omega} f(\bar{Y}|\bar{X},\bar{M}) f(\bar{X}|\bar{M}) \tag{2.33}$$

Here $Y$,$M$ and $X$ vectors contain the observed, known and latent RVs, respectively. The parameter space $\Omega$ indicates the entire probable domain of the latent variables which is

obtained from the range of the prior density function. Dealing with the discrete samples, solution of Eq. 2.33 is best to be sought through Monte Carlo integration:

$$f_{x_i} = [\prod_{j=1,\dots,K} \frac{1}{max(x_j) - min(x_j)}] \times \frac{\sum f(\bar{Y}|\bar{X},\bar{M}) f(\bar{X}|\bar{M})}{N} \tag{2.34}$$

Having calculated the evidence, the posterior likelihood is divided by this constant in order to be normalized:

$$f(\bar{X}|\bar{Y},\bar{M}) = \frac{f(\bar{Y}|\bar{X},\bar{M}) f(\bar{X}|\bar{M})}{E} \tag{2.35}$$

Later in this chapter the importance of evidence in model selection is also discussed.

### 2.3.4 Posterior PDF Estimation

The joint posterior PDF over the parameter space has been resulted from the previous packages. In the current package the focus is on reconstruction of the continuous posterior PDF for each single defined latent random variable in the network. According to subsection 2.2.6, this task is achieved in the following steps:

- Marginalization of the joint posterior PDF
- Systematic Resampling of the marginalized particle PDF
- Kernel Density Estimation on the resulted discrete resampled data

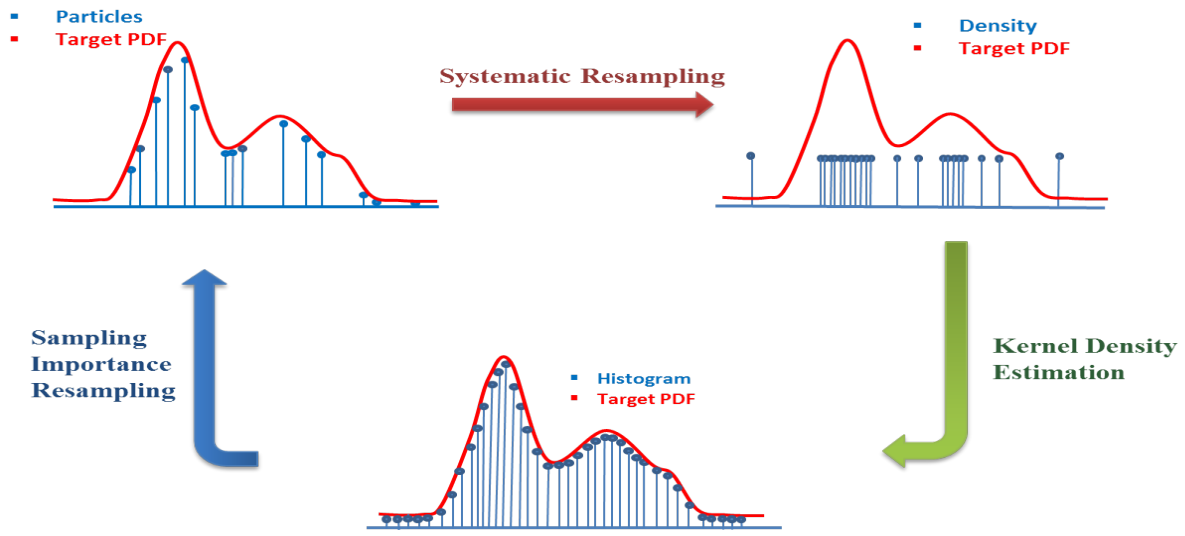



**Figure 2.14: Conversion between different representations of PDF; the particle representation is resulted from the SIR algorithm, converted to the density representation and finally estimated on the regular grid as a conventional "Histogram" representation which is more efficient in terms of number of stored values for representation of the PDF**

Fig. 2.14 depicts the three step conversion between different representations of the PDF.

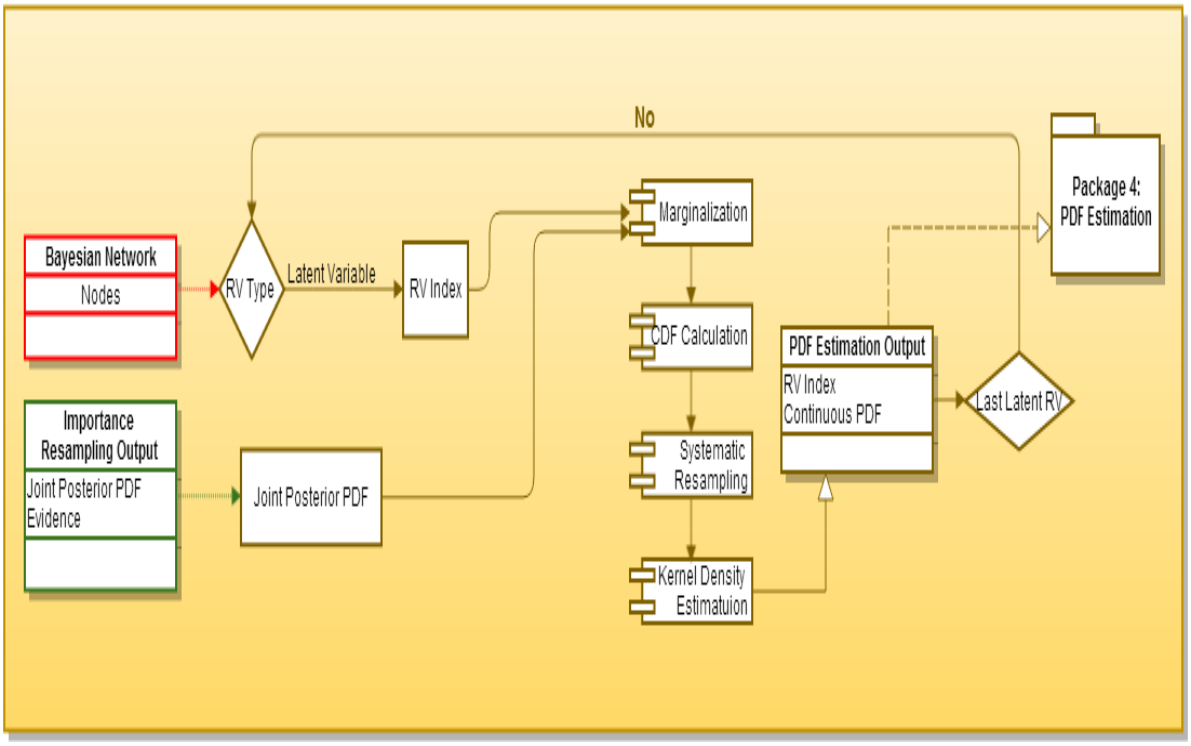



**Figure 2.15: Software package 4: PDF estimation; conversion from the joint posterior represented by random particles to the continuous PDF of each single parameter defined on a regular grid**

The optimum bandwidth of the kernel is calculated from Eq. 2.30. Fig. 2.15 summarizes this last package of the software. As a finalizing step, the outcome of the PDF estimation package for the slope parameter of the linear regression problem is presented in the next figure.

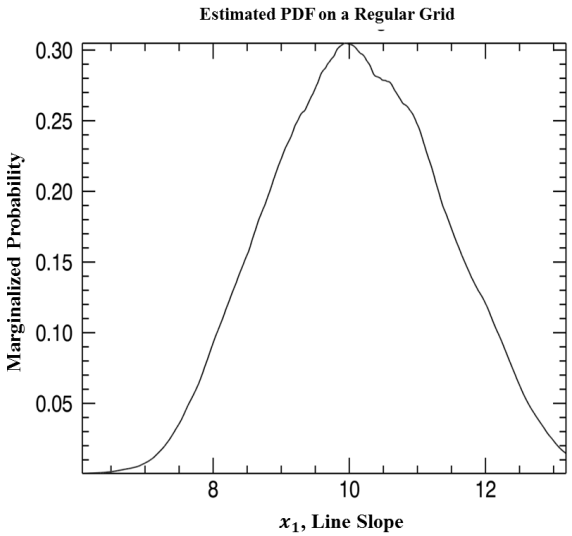



**Figure 2.16: Estimated PDF of slope parameter in linear regression example; the particle representation of the PDF is marginalized and converted to this continuous representation by the fourth package of the software**

### 2.3.5 Summary of The Software

Different packages of the software together with their composing modules have been reviewed in previous sections. Giving a summary of the software, Fig. 2.17 clarifies the interconnection between the results of the packages. The output of each package is the initial input to further packages, this relation is specified by the dashed lines in the figure. The explicit interconnection between the different subroutines in the modules as well as input and output of the subroutines are provided in the appendix.

It is worthwhile to mention that user interaction is only relevant to the first package of the software where the components of the probabilistic model are defined. The rest of the packages require information which are fed in from the lower level packages and have a full independent performance.

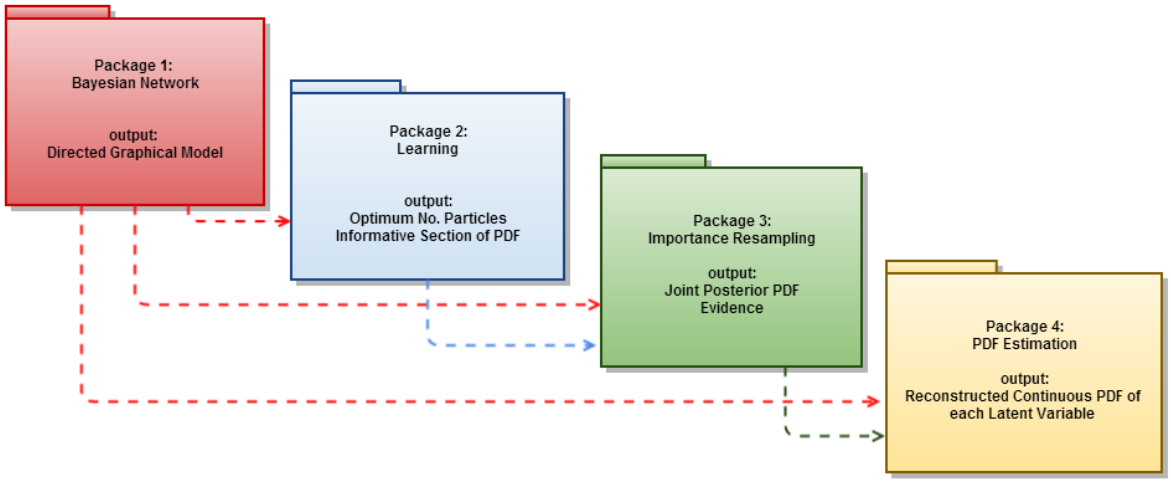



**Figure 2.17: Software data flow; the user information is required in the first package, the rest of the packages use information from their previous packages as well as the information included in the graphical model arranged by the first package**

The final output of the software is the evidence of the model as well as the imitated posterior PDF of each of the defined latent variables. This information is the basis for the two prominent inference tasks which are discussed in details in the coming section.

## 2.4 Inference on Bayesian Network

An elaborative explanation has been provided in the last section on deduction of the evidence as well as the posterior PDF of each single latent variable. The focus of this section is on fulfillment of two basic inference tasks, namely the parameter estimation and model selection, based on the deducted information.

### 2.4.1 Parameter Estimation

In an inverse problem, the presence of random noise in the measured data complicates the retrieval of the parameter. With finding the exact solution of the inverse problem being infeasible due to observations' random noise, the only choice is to estimate for them. The estimation process is thus an optimization problem to find the most appropriate parameters fitting the inverse problem.

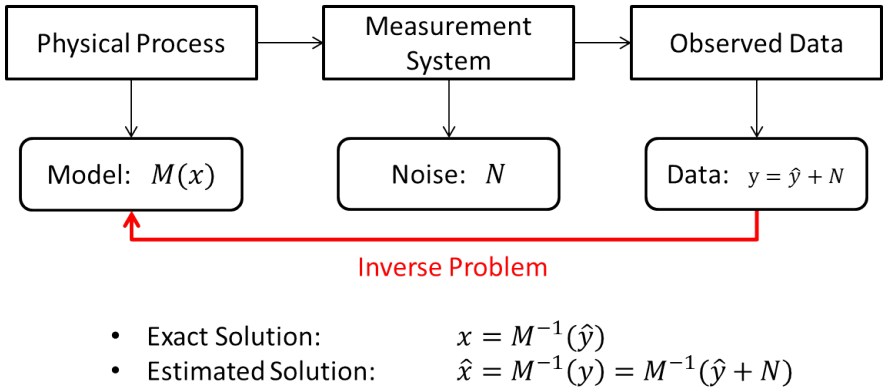



**Figure 2.18: Inverse problem and estimation as the solution; the random noise imposed by the measurement system complicates the modeling of the physical process behind the measured data**

Different criteria for the optimization problem open the vistas to different estimation methods. In a mathematical sense, the criteria for the optimization problem is minimization of a cost function for the estimation error, which gives rise to Minimum Mean Square Error (MMSE) estimators:

$$\hat{x}_{MMSE} = \text{argmin}_x\{E[(x - \hat{x})^2]\} \tag{2.36}$$

With $E$ defined as the expectation or the first moment of the squared error.

In a probabilistic sense, the optimization criterion is defined as maximization of the probabilities. The target probability function can be the likelihood of the data given the parameters which leads to the frequentist method of Maximum Likelihood Estimation (MLE):

$$\hat{x}_{MLE} = \text{argmax}_x\{f(Y|X, M)\} \tag{2.37}$$

An alternative to MLE is the Bayesian approach that provides the complete picture of the probabilistic model by introducing the prior knowledge about the parameters to the estimation scheme. The Bayesian estimation is based on the Bayes' theorem and attempts

to maximize the posterior likelihood to find the Maximum A Posteriori (MAP) estimation of the parameters:

$$\hat{x}_{MAP} = \text{argmax}_x\{\frac{f(Y|X,M)f(X|M)}{f(Y|M)}\} = \text{argmax}_x\{f(X|Y,M)\} \tag{2.38}$$

The presence of prior probability discriminates the MLE and MAP estimators. Naturally, in presence of a non-informative prior, as a uniform PDF, the two estimators are equivalent.

Having the complete marginalized posterior PDF of each single parameter from the software, the MAP estimate is simply found by the maximum of this PDF. In order to boost the accuracy of peak retrieval and to have an assessment of the precision of estimated value, a Gaussian curve is fitted around the maximum of the histogram. The mean and standard deviation of the fitted Gaussian function indicate the MAP estimate of the parameter and the estimation precision, respectively.

$$\hat{x}_{MAP} = \mu$$
$$\hat{\sigma}_x = \sigma$$

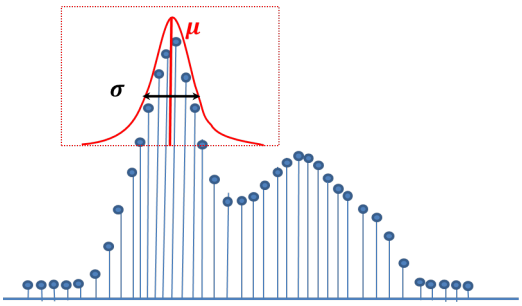



**Figure 2.19: MAP estimation from retrieved PDF of a single parameter; the fitted Gaussian curve to the peak of the PDF gives the MAP estimate and the precision of the estimation**

## 2.4.2 Model Selection

Up to this point, only one model was considered for evaluation of the data. In practical modeling problems, a finite set of models is proposed to retrieve the information from the data. The model selection is regarded as the methods for decision on the optimum model which can best retrieve the information.

The proposed models could be from the same family with different number of parameters, such as determination of number of significant parameters in linear regression or polynomial fit problems or completely different types of models.

To explain the methods of the model selection, assume the simple linear regression problem:

$$Y = x_0 + mx_1 + m^2x_2 + \ldots + m^kx_k \tag{2.39}$$

The question which model selection answers in this case is: what is the optimum $K$ which can best describe the underlying information behind the observed data. Two contradictory criteria complicate the answer:

1. The accuracy of the model fit to the data, the quantity which is referred to as *goodness of fit* and can be interpreted as the likelihood of the data given the model i.e. $f(Y|X(K), K)$
2. The model complexity which can be translated into the number of parameters that describe the model, here denoted by $K$

Increasing the model complexity leads to better goodness of fit of the model. The reason is that the excessive parameters in the model give it the flexibility to describe random features of the data, which should be attributed as the noise instead of the informative underlying relation among the data. This problem which is regarded as the *over-fitting* or *over-parameterization* must be avoided by a suitable model selection criterion.

To set a compromise between the model complexity and the goodness of fit, the likelihood is *penalized* for the complexity of the model. The *penalized likelihood* model selection methods are generally formulated as:

$$\hat{M}(\hat{K}) = \text{argmax}_{X(K)}\{\ln(f(Y|X(K), M(K))) - C(M(K))\} \tag{2.40}$$

The term $C(M(K))$ is the *penalization term* also known as the Ockham's factor. A number of criteria can be found in the literature that follows the penalized likelihood scheme. These methods are mainly distinguished by the defined penalization term. Common examples of these criteria are: Bayesian Information Criterion (BIC) [20], Akaike Information Criterion (AIC) [21, 22], Minimum Description Length (MDL) [23, 24] etc. detailed explanation of which are beyond the scope of this section.

The alternative approach to the penalized likelihood is the Bayesian model comparison [7, 10, 25]. This method is based on the evidence of the model and introduces the prior probabilities into the decision frame:

$$\begin{aligned} \hat{M}(\hat{K}) &= \text{argmax}_{M(K)}\{(f(Y|M(K)))\}, \\ \hat{M}(\hat{K}) &= \text{argmax}_{M(K)}\{\int_{\Omega} f(Y|X(K), M(K))f(X(K)|M(K))\} \end{aligned} \tag{2.41}$$

Unlike the previously defined criteria which have an explicit complexity penalty, the Ockham effect here is more subtle: The evidence has the penalization of the model complexity implied in its structure. A complex model increases the likelihood probability in one hand, while on the other hand it decreases the prior probability of the parameters since the prior is spread in wider area of the parameter space. In this situation, unless a highly informative prior is considered, as the dimension of parameter space increases, a smaller proportion of the prior's probability mass falls within the region of parameter space in which the likelihood is significant. Hence the value of the evidence falls and models with more parameters are penalized [26].

The choice of parameter priors for Bayesian model comparison is a challenge: if the priors are too diffuse, the penalization term is exaggerated and the simplest model is always

chosen. Therefore it is important to choose a proper prior to the model. The sensitivity of the model selection to this choice is discussed in [25]. A variety of strategies are introduced to choose the prior effectively examples can be found in [25, 27, 28]. In the proposed software, the problem of non-informative priors is handled via the importance resampling in the SIR. The resampling step helps in concentration of the joint posterior PDF on the informative part of the prior, thus preventing the unnecessary diffusion of the initially defined prior from degrading the evidence.

# Chapter 3

# Demonstrative Applications in Persistent Scatterer Interferometry

Two applications have been considered and tested in the PSI framework to demonstrate the performance of the developed software in different inference tasks. Prior to describing the applications, a short overview of the PSI is provided in the first section; with the motivation of the two applications reviewed at the end of this section. The SAR signal statistics are then reviewed in the second section as a required background information. Each of the applications is then formulated in a probabilistic model and treated with the Bayesian inference. The performance of the algorithms is evaluated using synthesized data and finally the methods tested on TerraSAR-X data.

## 3.1 Introduction to Persistent Scatterer Interferometry

Invented at Politecnico di Milano (POLIMI) in the late 1990$s$, the PSI is at the moment the most practical approach in retrieval of geodynamic signal from the time series of radar interferometric phase with accuracy in the regimes of millimeter [3]. From the advent of PSI to date, a number of scientific teams around the world have developed practical frameworks based on the same initial idea.

To clarify the strength of the PSI over other conventional InSAR methods, it is necessary to start from the characteristics of the interferometric phase. Radar interferometric phase is caused due to the difference in measured distances of a scatterer on the ground to the radar sensors in two different acquisitions. This difference is related to the topography of the scatterer [1]. The change in the topography can be monitored by time series analysis of the interferometric phase. However, the superposition of other temporal and spatial signals on the topographic signature complicates this analysis. The interferometric phase is comprised of the following contributions:[29]

$$\phi_{ik} = \phi_{topo} + \phi_{defo} + \phi_{scat} + \phi_{orbit} + \phi_{atm} + \phi_{noise} + N \cdot 2\pi \tag{3.1}$$

- $\phi_{ik}$: The interferometric phase obtained from the interferograms. Interferogram is the phase difference of a reference complex image ($u_i$), called *master* and a secondary

complex image ($u_j$), known as *slave* that is acquired at a different time and/or from a slightly different orbital position:

$$\phi_{ik} = u_i \cdot u_k^* \tag{3.2}$$

- $\phi_{topo}$: The phase induced by terrain topography. The height to phase conversion factor ($\beta$) links the topographic height to the corresponding part of the phase:

$$\phi_{topo} = \frac{-4\pi}{\lambda} \cdot \frac{B_{prep}}{R. \sin\theta} \cdot h = \beta \cdot h \tag{3.3}$$

  In here, $B_{prep}$ is the perpendicular baseline between the two acquisitions, $R$, the master slant range and $\theta$ is the radar incidence angle at master position.

- $\phi_{defo}$: The phase induced by terrain deformation in time interval between the master and slave acquisitions. Deformation imposes range difference ($\Delta r$) in the radar line of sight direction which is linked to the interferometric phase by the following relation:

$$\phi_{defo} = \frac{-4\pi}{\lambda} \cdot \Delta r \tag{3.4}$$

  The deformation signal is decomposed into a linear combination of basis functions of the form:

$$\Delta r = \sum_{d=1}^{D} \alpha_d \cdot p_d(t) \tag{3.5}$$

  In which $p_d(t)$ is the polynomial of order $D$ as a function of time, and $\alpha_d$ is the amplitude of this polynomial. Although in general any basis function may be considered in deformation modeling, usually a linear deformation model is considered:

$$\Delta r = \alpha(x) \cdot T = v_{LOS} \cdot T \tag{3.6}$$

  In here, $T$ is the temporal baseline, and velocity of deformation in the line of sight ($v_{LOS}$) is the target of estimation.

- $\phi_{scat}$: The object scattering phase related to the path length traveled in the resolution cell. This component is the cause of temporal and geometrical decorrelation in interferograms

- $\phi_{orbit}$: The phase induced by inaccuracies in the orbit determination of the sensor

- $\phi_{atm}$: The atmospheric phase accounting for the signal delays caused in propagation path through the troposphere and partially through the ionosphere [2]

- $\phi_{noise}$: The additive phase noise

- $N \cdot 2\pi$: The ambiguous integer cycles of phase which are not contained in the observation and retrieval of which is in the context of phase unwrapping [1]

The temporal and geometrical decorrelation as well as the atmospheric induced signal path variations are the main limits in conventional interferometric approaches in retrieval of deformation signals [6]. The PSI technique overcomes these limits by two distinguished strategies: [3, 29]

1. Restriction of interferometric analysis to the coherent radar scatterers with maximum phase stability to eliminate the effect of geometrical and temporal decorrelation
2. Consideration of the atmospheric induced signal in the observation model in order to separate this error source from the interferometric phase

In the gist of it the PSI is a spatiotemporal analysis. In the temporal dimension, interferometric phase is obtained by interferogram formation between all available acquisitions and a single master scene. As opposed to the conventional interferometric methods, quality of the interferograms is not compromised by the single master strategy. This is because merely a sparse distribution of point wise coherent scatterers, known as PS, is considered in the analysis, for which the phase behavior is stable regardless of the temporal and spatial baselines between the acquisitions. The achieved interferograms are later refined by subtraction of the topographic phase modeled from a coarse Digital Elevation Model (DEM). Result is the *differential interferograms* from which the major part of the topographic signal is eliminated. The residual topographic phase is later modeled together with the deformation signal parameters.

In the spatial dimension, the difference between the differential interferometric phases of the nearby PS candidates further decreases the undesired signals [30]. Specifically the atmospheric signal is considerably reduced in this manner since it is spatially correlated in the distance of approximately 1 kilometer [31]. The resulted spatiotemporal differential phase is the initial observation for retrieving the deformation signal.

Having this broad overview of PSI, a more detailed explanation of the framework is provided based on the German Aerospace Center (DLR)- developed PSI module integrated in the GENEric System for Interferometric SAR (GENESIS). PSI-GENESIS module is proved to be a robust processing chain and has been chosen as the reference for validation of products of different PSI operational service providers [32].

### 3.1.1 The Persistent Scatterer Interferometry Framework

The aim here is to give a short review of different processing levels of the DLR's PSI-GENESIS processing chain. An elaborative explanation of each step is provided in [29]. The general steps are summarized in Fig. 3.1 with the explanations coming in the following:

***Interferogram formation***

As mentioned, PSI interferometric stacking is carried out using a single master scene. The master scene is selected by maximizing the expected stack coherence. The stack coherence is defined based on the temporal and spatial baselines and the mean *Doppler centroid frequency* difference between the pairs of master and slaves [29]. After selection

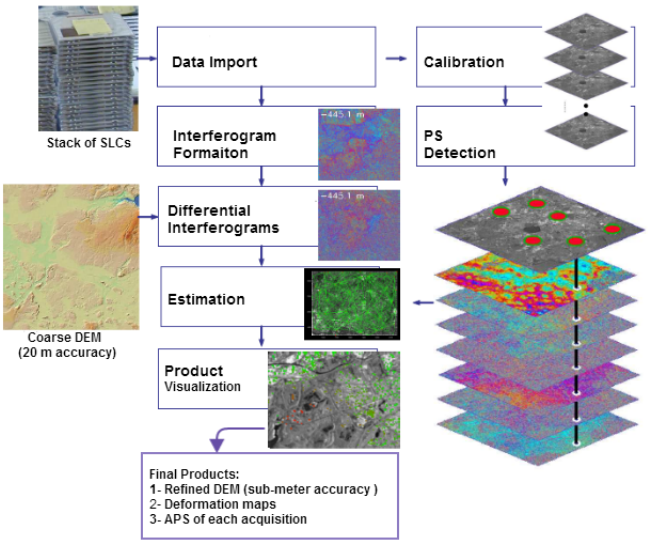



**Figure 3.1: Overview of the PSI-GENESIS persistent scatterer interferometry framework [modified from Adam et. al., 2004]**

of the master scene, all Single Look Complex (SLC)s are coregistered to it. The coregistered SLCs are then resampled to and paired with the master to form the interferograms.

### *Differential Interferograms*

At this step the observation geometry of radar acquisition is simulated. The precise orbit parameters as well as a coarse DEM are used to simulate the topographic phase between the interferometric pairs. The simulated phase is then eliminated from the interferometric phase to form the differential interferograms.

Even a highly accurate DEM does not guarantee the complete removal of the topographic phase as the phase centers of the PS points are not known in advance. The residual topographic phase is later considered in the estimation step and provides the refined DEM. Thus, the accuracy of the DEM is not critical.

### *Detection of Persistent Scatterers*

From all the pixels of the differential interferograms, deformation analysis is restricted to persistent scatters i.e. the sparse points with long time stable back scattering behavior such as man-made structures. The detection of PSs is possible via time series analysis on the amplitude of the SLCs, since the amplitude is almost insensitive to most of the phenomena contributing to the phase content [3]. The quantity that can be inferred from the amplitude time series and be related to the stability of the interferometric stack is the SCR of the radar scatterers. The relation between phase dispersion and the SCR is known to be:[30]

$$\sigma_\phi \simeq \frac{1}{\sqrt{2.SCR}} \tag{3.7}$$

Two methods have been proposed for estimation of the SCR:

1. Amplitude Dispersion index: is defined as the ratio of standard deviation of the amplitude stack to mean of the stack and estimates the temporal SCR by: [3]

$$D_A = \frac{\sigma_A}{\mu_A}; \qquad \hat{SCR} = \frac{1}{2.D_A^2} \tag{3.8}$$

2. Spatial SCR: uses certain spatial estimation window on the amplitude of each single SLC and estimates the signal and clutter power of the scatterer inside the window. This estimation is valid under the assumption that the power of clutter around a pixel is equal to the power of clutter inside the resolution cell.[33]

Both methods are unbiased estimators of phase dispersion in the high SCR levels, but when it comes to low SCR levels both underestimate the phase error [30].

Based on the estimated SCR, the PS candidates are detected by applying threshold on the SCR values. Reasonable threshold is $SCR = 2$ which allows the phase error of less than 0.5 radians for the detected PS candidates.

***Estimation***

Up to this point the differential phase of persistent scatterers are at hand for commencement of the deformation analysis. The differential phase functional model is given by:

$$\phi_{DInSAR} = \phi_{topo_{resi}} + \phi_{defo} + \phi_{atmo} + \phi_{orbit} + \phi_{noise} \tag{3.9}$$

In order to further decrease the disturbing signals, the phase difference between the nearby PS points are considered. Forming this double difference observation, the phase contributions from the orbit and atmospheric effects are eliminated from the differential interferograms. The resulted relative differential phase is modeled as:

$$\Delta\phi_{DInSAR} = \Delta\phi_{topo_{resi}} + \Delta\phi_{defo} + \Delta\phi_{noise} \tag{3.10}$$

This observation allows for relative estimation of the deformation parameters as well as the residual topography.

The estimation is carried out in two levels. In the preliminary level, a reference network of selective PS candidates with high SCR are chosen in a coarse grid of 1 km. The spatiotemporal processing is performed on this reference network. The arcs between the selective candidates are formed and the relative parameters are sought by Least squares AMBiguity Decorrelation Adjustment (LAMBDA) method in the temporal dimension. The estimated relative parameters are then integrated by a spatial integration based on least squares adjustment to provide global parameters on each point of the network. A reference point must be introduced in the network as the starting point of the integration, the choice of which is critical in the final results of estimation. Usually this point is chosen in an a priori known geodynamical stable region where the deformation is expected to be zero. Finally the residuals of the network are exploited in an outlier detection process in order to eliminate the points with inconsistent estimation results.

In the final estimation, the remaining PS points are linked to their closest point from the reference network by a single arc, with their parameters being sought by LAMBDA estimator. Having the absolute parameters of reference points from the preliminary level, the parameters of the current level PSs are derived through the simple integration between the two points.

Finally the residual phase after elimination of estimated topography and deformation are exploited to estimate the Atmospheric Phase Screen (APS). The residual phase contains the atmospheric and the un-modeled deformation components as well as a white noise. The atmospheric effects are correlated in space and uncorrelated in time, while the residual deformation is correlated in time only. Therefore, a spatial low pass filtering separates the atmospheric phase and a temporal high pass results in separation of the residual deformations. After retrieving the atmospheric phase on the sparse PS points, the APS is estimated by interpolation of these values on a regular grid.

### 3.1.2 Potential Improvements in the Framework

From its advent to date, PSI has been the target of studies for many scientific teams and a wide range of improvements in the algorithm have been proposed so far. The attempt here is made to introduce and investigate two specific areas which can be improved in the framework and to later treat them with the developed Bayesian inference software of Chapter 2. A short introduction on these two cases is provided in the following.

Importance of coherence of PS points has been justified to this point; still implicit is the effect of density of these points on the final products of PSI. The PS points are regarded as the samples for the final PSI products, e.g. APS and deformation map, thus the higher the spatial frequency of these samples is, the more reliable the recovered signals will be. A trade off appears between the density of the PSs and their coherence, rendering the detection of the PSs a critical task. On the one hand the false detection of low coherent points impose vast computational burden in post process outlier detection methods to find and discard them, on the other hand discarding the coherent points, as a result of missed detection, decreases the density of the PSs and leads to information loss. It is the task of detection criterion to set a balance between these two. Improving this criterion can therefore lead to integration of more stable points in one hand or reducing the post processing effort for removal of miss detected points on the other. As discussed in previous section, SCR is the base criterion in detection. However, the two introduced methods of SCR estimation, i.e. the amplitude dispersion index and the spatial SCR estimator, are both biased and underestimate the phase error [30]. In the latter, bias stems from the fact that the surrounding pixels in spatial estimation window are not perfect representatives for the clutter [29], while in the former it is due to simplification of statistics of the amplitude stack [3, 34]. Alternative to these two methods, the exact signal statistics can be considered to estimate the temporal SCR. Formulation and evaluation of this alternative approach is expanded in section 3.3 as the first application of the Bayesian inference software.

As previously discussed, PSI is restricted to the pixels with phase coherence over the entire time series. Although initially vital to the PSI framework, this restriction results in loss of information in regions where presence of PSs is limited. Relaxing this constraint,

from stability in the entire interferometric stack to partial stability in a sub-set of stack, is the second area of focus in this thesis. Inclusion of partially stable points referred to as Temporal Persistent Scatterer (TPS), in the deformation analysis increases the density of the PS points and provide new source of information for deformation monitoring [35]. Inference on such scatterers is in the scope of section 3.4.

Both inferences are based on the calibrated amplitude stack, therefore prior to the explanation of these applications an introduction to SAR amplitude statistics is provided in section 3.2.

## 3.2 SAR Coherent Imaging Statistics

Each pixel in a focused complex SAR image represents the superposition of back scattered signals from all scattering elements falling into the corresponding resolution cell on the ground, in other words:

$$z = x + jy = \sum_{i=1}^{M} a_i . \exp(j\phi_i) = \sum_{i=1}^{M} a_i \cos\phi_i + j \sum_{i=1}^{M} a_i \sin\phi_i \tag{3.11}$$

With $z$ as a complex value, $x$ and $y$ as real numbers and $j$ as the imaginary unit of the complex value, $i$ as the index of the sub-scatterers and finally, $\phi_i$ and $a_i$ as the amplitude and phase of the back-scattered signal from each sub-scatterer, respectively. Fig. 3.2.I visualizes this equation.

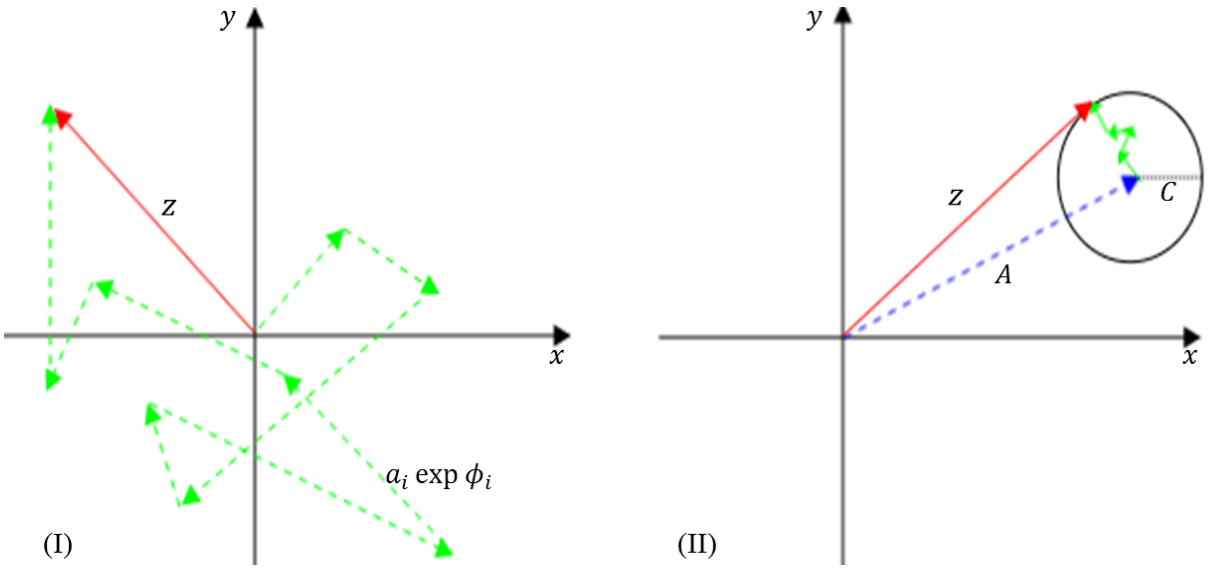



**Figure 3.2: Observed complex SAR signal model for (I) Distributed scattering occurred in presence of multiple sub-scatterers inside the resolution cell and (II) Point scattering occurred in presence of one dominant scatterer (A) and multiple minor background sub-scatterers (C)**

In presence of sufficiently large number of sub-scatterers the central limit theorem applies and the resulted complex value follows a circular Gaussian process [1], with its probability distribution function given as followed:

$$f(z|\bar{I}) = \frac{1}{\pi \bar{I}} \exp\{-\frac{x^2 + y^2}{\bar{I}}\} \tag{3.12}$$

where, $\bar{I}$ is defined as the expected intensity of the pixel i.e. $\bar{I} = E\{I\} = E\{(x^2 + y^2)\}$. In order to derive the amplitude statistics from the circular Gaussian process, two extreme scattering scenarios are considered: the distributed scattering versus point scattering; represented in part (I) and (II) of Fig. 3.2, respectively. The latter is decomposed into a sufficiently high number of random sub-scatterers within a resolution cell with no single sub-scatterer remarkably dominating the others. Most natural scatterers such as forests, agricultural fields, rough water, soil etc. fall into this category for space-borne

SAR systems. The amplitude of this type of scatterers yield the one-parameter *Rayleigh* distribution function: [1]

$$f(A|\sigma) = \left(\frac{A}{\sigma^2}\right) \exp\left(-\frac{A^2}{2\sigma^2}\right) \tag{3.13}$$

with amplitude defined as $A = \sqrt{I} = \sqrt{z^2}$ and expected signal intensity of: $\bar{I} = E\{I\} = 2\sigma^2$.

The second scattering scenario addresses the presence of one dominant point scatterer in a background of several minor sub-scatterers, as it can be found in artificial objects and man-made structures of the urban areas. The amplitude in this case is described by the two-parameter Rice distribution: [3]

$$f(A|\nu,\sigma) = (\frac{A}{\sigma^2}) \exp(-\frac{A^2+\nu^2}{2\sigma^2}) I_0(\frac{A\nu}{\sigma^2}) \tag{3.14}$$

where, $\nu^2$ and $\sigma^2$ represent the intensity of the dominant scatterer and the background clutter, respectively and $I_0(.)$ is defined as the modified Bessel function of first kind and zero order: [36]

$$I_0(x) = \frac{1}{\pi} \int_0^{\pi} \exp(j.x.\cos\theta)\, d\theta \tag{3.15}$$

Shown in Fig. 3.3, the two introduced amplitude models are the basis for the two presented applications in the following.

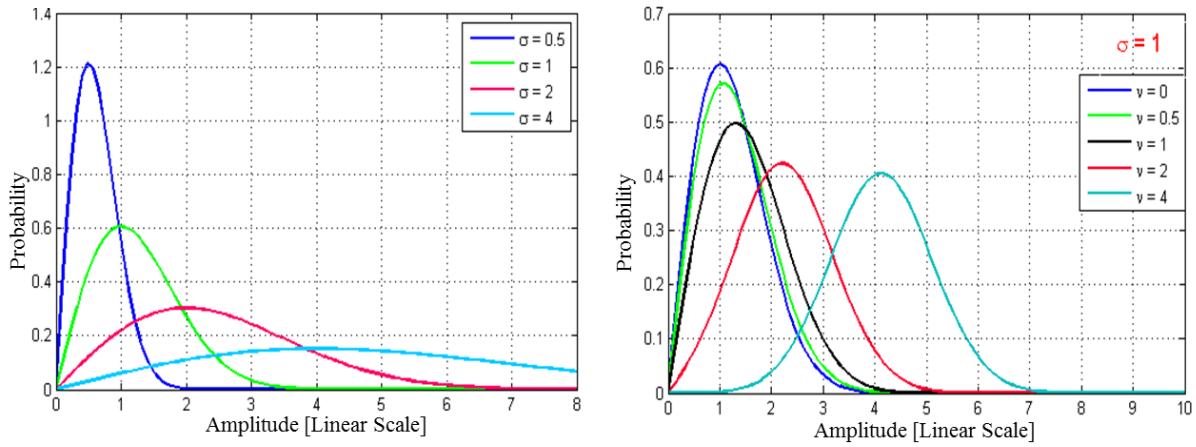



**Figure 3.3: PDF of the two introduced amplitude models, (I) Rayleigh PDF with different $\sigma$ values (II) Rice PDF with fixed $\sigma$ and variant $\nu$**

## 3.3 Detection of Persistent Scatterers

In this section the amplitude time series of the coregistered SAR images is analyzed to infer information about the scattering mechanism occurred in each resolution cell and assess the coherence of the scatterers. Considering the two models of persistent and distributed scatterers, the Bayesian model selection helps to decide which scattering model fits the observed amplitudes. The parameters of the best fit model are then estimated using the Bayesian approach. Having the model parameters estimated, the SCR is obtained and finally used as an estimation of phase dispersion. Applying a threshold on the phase dispersion the persistent scatterers are selected. These selected PS candidates are the target for further phase time series analysis of the coherent advanced InSAR techniques.

The detection method based on the Bayesian approach is described in the first section. In sub-section 3.3.2 relation between the estimated SCR and phase coherence is assessed and compared to alternative method of amplitude dispersion index using the simulated data. Finally the established and evaluated method is applied to TerraSAR-X data stack in sub-section 3.3.3 to test the feasibility of this approach in dealing with real data.

### 3.3.1 Problem Statement and Solution with Bayesian Inference

Having the stack of coregistered SLCs, the amplitude time series is used on a pixel-by-pixel basis to perform model selection between two defined extreme scattering models of Rice and Rayleigh. The Bayesian network corresponding to each of the models is provided in Fig. 3.4.

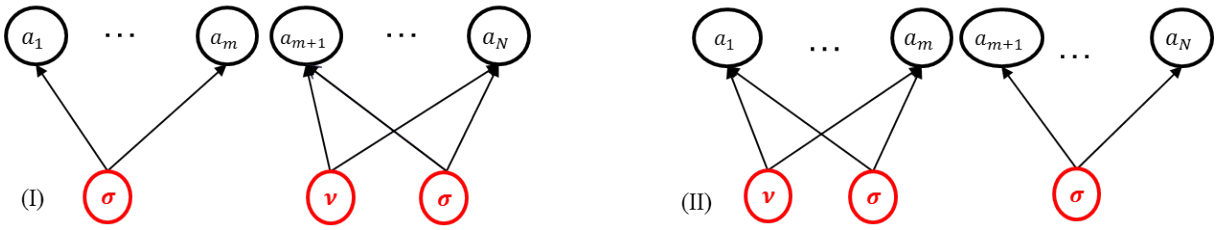



**Figure 3.4: Bayesian network of two amplitude models; (I) Rayleigh model with one parameter appropriate for distributed scatterers and (II) Rice model with two parameters proposed for point scatterers**

The black nodes in both networks correspond to the observed amplitude while the red ones indicate the model parameters. Table 3.1 and 3.2 summarize the characteristics of assigned PDFs to each type of nodes in the two given networks. In this table $A$ indicates the vector containing the amplitude stack and *stddev* is the standard deviation operator.

**Table 3.1: Node characteristics of the Rayleigh Model**

| Node | Node PDF | Upper Bound | Lower Bound |
|---|---|---|---|
| $A$ | Rayleigh | ... | .... |
| $\sigma$ | Uniform | $\max\{\bar{A}\} + \text{stddev}\{\bar{A}\}$ | $\min\{\min\{\bar{A}\} - \text{stddev}\{\bar{A}\}, 0\}$ |

**Table 3.2: Node characteristics of the Rice Model**

| Node | Node PDF | Upper Bound | Lower Bound |
|---|---|---|---|
| $A$ | Rice | ... | .... |
| $\nu$ | Uniform | $\max\{\bar{A}\} + \text{stddev}\{\bar{A}\}$ | $\min\{\min\{\bar{A}\} - \text{stddev}\{\bar{A}\}, 0\}$ |
| $\sigma$ | Uniform | $\max\{\bar{A}\} + \text{stddev}\{\bar{A}\}$ | 0 |

Taking each of the models in turn, Bayesian inference is implemented on the amplitude stacks in a pixel-by-pixel basis. The evidence is calculated for each of the models and the model with higher evidence is chosen as the more appropriate fit to the observed amplitude time series. Finally the parameters of the chosen model are estimated by Bayesian approach. These parameters are the expected signal amplitude for Rayleigh model, i.e.:

$$\hat{A} = E\{A\} = \sqrt{2\sigma^2} \tag{3.16}$$

as well as the expected signal and clutter amplitude in case of Rice model:

$$\hat{A} = E\{A\} = \nu \tag{3.17}$$

$$\hat{C} = E\{C\} = \sqrt{2\sigma^2} \tag{3.18}$$

Having the signal and clutter amplitude estimated for the Rice detected pixels, it is now possible to calculate the SCR for such pixels.

$$\hat{SCR} = \left(\frac{\hat{A}}{\hat{C}}\right)^2 \tag{3.19}$$

Finally detection of PSs is possible by applying threshold on the estimated SCR values.

Although a number of alternative approximation methods are proposed for estimation of the signal parameters [37], there are two basic advantages in the Bayesian approach over the analytical approximations:

1. The mathematical complexity as well as computational iterations involved in the approximation methods are by far more expensive than the proposed Bayesian method

2. Dependent on the complexity of the sub-resolution scattering structure, the marginal PDF of the dominant signal may be multimodal. In such cases, approximation methods may falsely fall into the local maximum. On the contrary, having the complete PDF of the amplitude reconstructed, the Bayesian approach guarantees to find the global solution to the problem.

### 3.3.2 Investigation with Synthetic Data

The aim here is to have a performance assessment on the proposed Bayesian SCR estimator on the one hand, and to investigate the relation between the estimated SCR and the phase stability on the other hand. For this purpose, a stack of $N = 50$ complex-valued

observations is synthesized. Assuming a constant coherent signal in decorrelating clutter, the observations are simulated as:[34]

$$z_i = s + n_i = s + (x_i + j \cdot y_i); \quad i = [1, \ldots, N] \tag{3.20}$$

In which $z$ simulates the recorded SAR complex valued signal in presence of a significant signal from a dominant scatterer. The complex value is comprised of $s$ as the coherent signal and $n_i$ as the decorrelating clutter, the real $(x_i)$ and imaginary $(j \cdot y_i)$ part of which follow a circular Gaussian random process with standard deviation $\sigma_n$. In this case the signal and clutter power are derived as $A = |s|^2$ and $C = 2\sigma_n^2$, respectively. For the experiments, the signal level is set to 1 and the noise standard deviation is varied in range $[0.05, 0.75]$ in steps of 0.05. For each variation the trial is repeated 400 times. In each trial, the true phase noise is empirically calculated as the standard deviation of the phase of the complex stack:

$$\sigma_\phi = [\frac{1}{N} \sum_{i=1}^{N} (\phi_i - \mu_\phi)^2]^{1/2} \tag{3.21}$$

Where $\phi_i$ is the argument of the complex values and $\mu_\phi$ is the mean of the arguments in the stack. The SCR is estimated by two introduced methods of Bayesian estimation followed from Eq. 3.19 and the amplitude dispersion followed from Eq. 3.8. Having the SCR, the phase noise is estimated using Eq. 3.7. Fig. 3.5 summarizes the result of this experiment. Revealing lower bias in low SCR values, the Bayesian method is a better approximation of the phase error compared to the amplitude dispersion index.

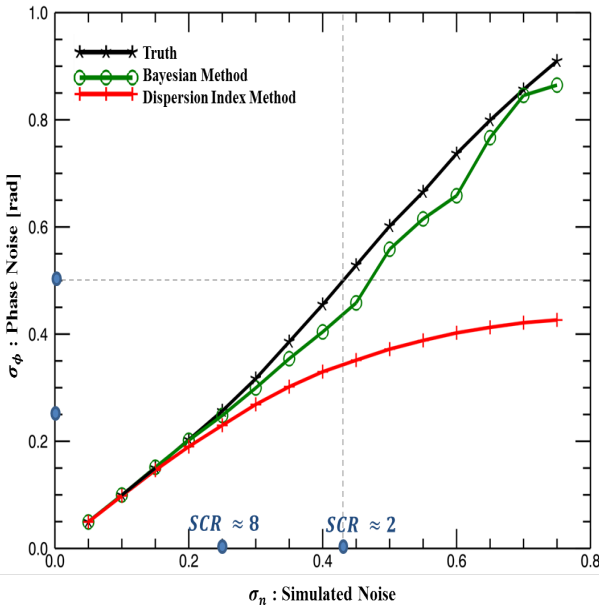



**Figure 3.5: Comparison between the estimated phase error from the amplitude time series analysis; the two methods of amplitude dispersion index and Bayesian estimation are compared against the true phase, the grey line indicates the common threshold for PS detection (SCR = 2)**

The scatter plot of the estimated versus the true phase noise depicts the result of all trials in the experiment (Fig. 3.6). The plot reveals that precision of both methods decrease with phase noise; in other words, both estimators are sensitive to noise in low SCR. This sensitivity is more pronounced in the Bayesian approach. It is also deduced that the amplitude dispersion method is more precise while the Bayesian method is more accurate in estimation of phase variations.

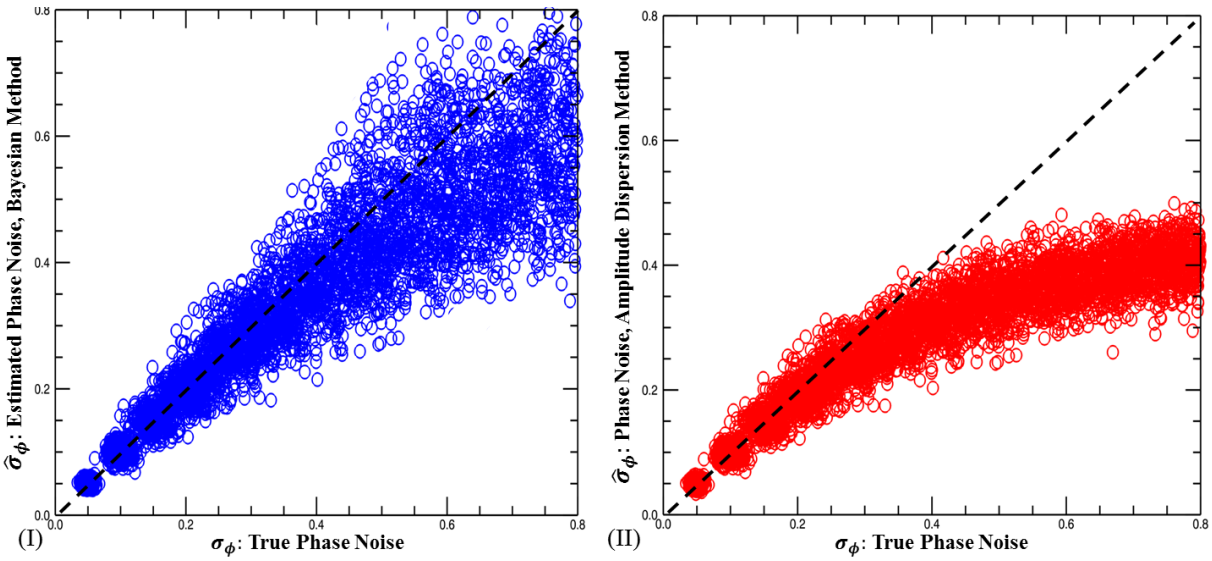



**Figure 3.6: Scatter plot of the estimated vs. simulated phase noise using two methods of (I) Bayesian estimation and (II) amplitude dispersion index; the dashed line represents unbiased estimation, the Bayesian method is more accurate but less precise compared to the amplitude dispersion index method**

Having evaluated the method, it will be further tested on the real data in the following section.

### 3.3.3 Experiment on Real Data

A small test site over the Hohenzollern bridge of Cologne city in Germany is chosen in this section to investigate the proposed method (Fig. 3.7). The dataset is comprised of 53 high resolution spotlight TerraSAR-X SLCs acquired in time span of March 2009 to December 2010.

The different scattering phenomena of the test site, ranging from persistent scatterers in the metallic parts of the bridge to the decorrelating distributed scatterers on the side walk, make it an interesting region for the experiment.

In the first step of the investigation, model selection is implemented on the stack of calibrated amplitude in a pixel-by-pixel basis. The SCR is then estimated in the second step for the Rice classified pixels. Fig. 3.8 is dedicated to the obtained results with (I) showing the classification map and (II) reporting the estimated SCR. As expected, the metallic parts of the bridge are mostly detected as Rician scatterers with high SCR while

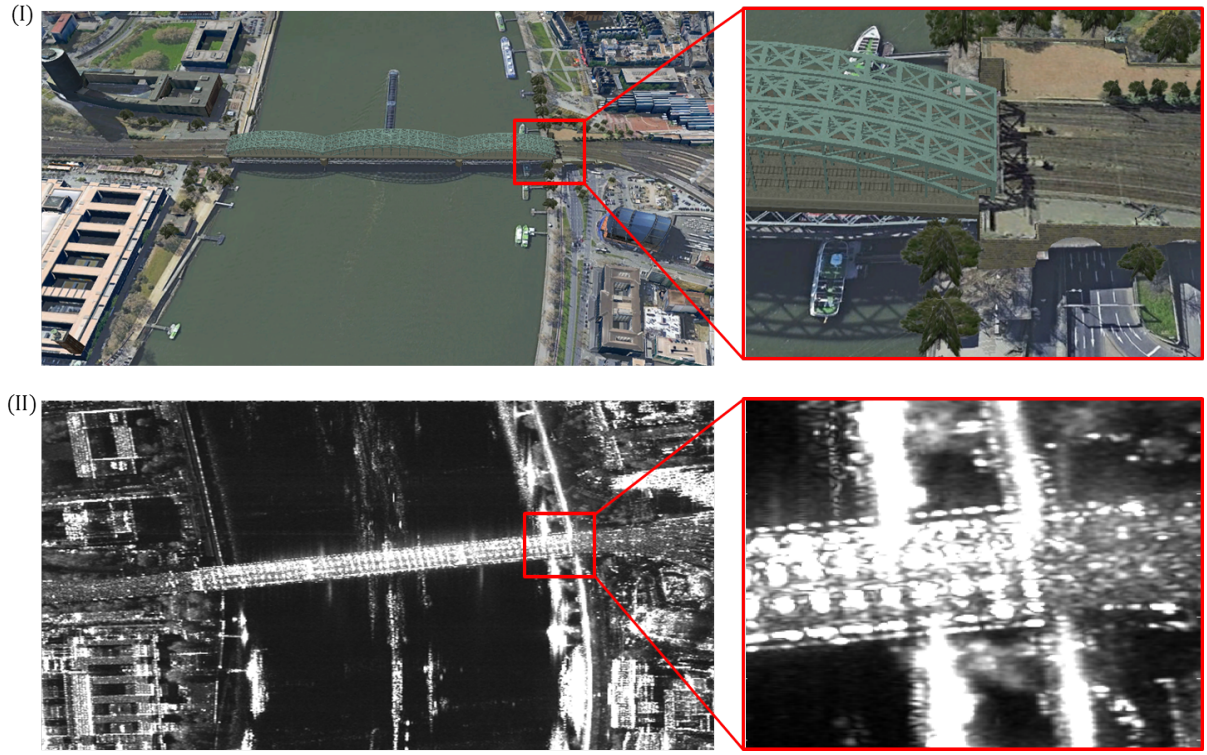



**Figure 3.7: Test site 1, the Hohenzollern bridge, Cologne-Germany. (I) Optical image of the site obtained from Google Earth, (II) calibrated multi-looked spotlight TerraSAR-X amplitude image of 100 SLCs over the test site**

the side walk and water parts are rejected for the Rician model.

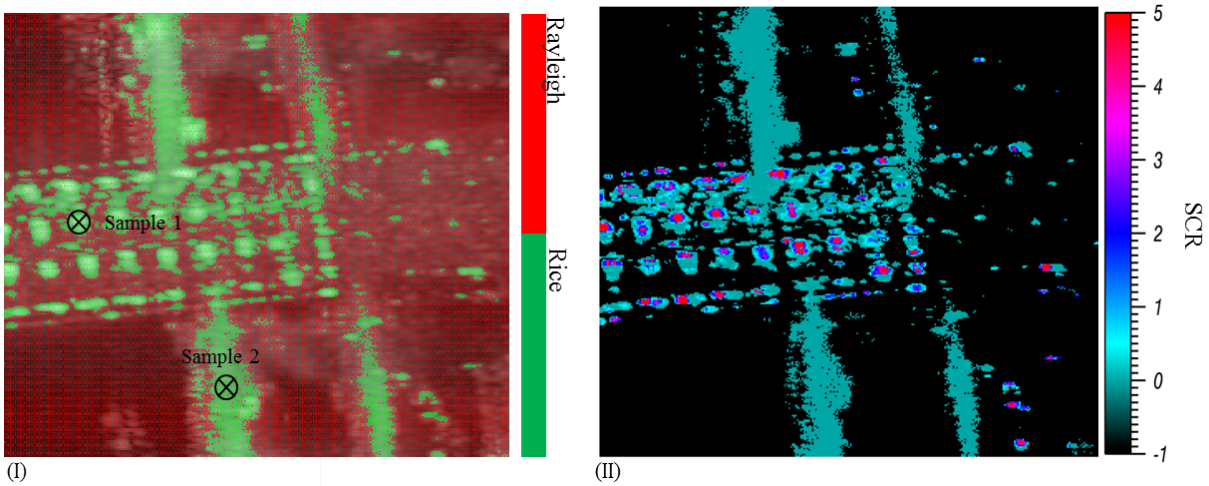



**Figure 3.8: Inference on first test site; (I) Model selection result, classification between Rice and Rayleigh models, the results obtained for the two samples indicated by cross symbols are studied later (II) Estimation result, SCR of the Rayleigh scatterers is set to -1**

To further inspect the result, two random samples are chosen from the scene: the first on the metallic bridge, as a perfect representative of Rice-distributed scatterers (Fig. 3.9) and the second on the edge between the water and side walk which is initially classified as Rician but finally is estimated to have a very low SCR (Fig. 3.10). These two samples

are marked in Fig. 3.8.

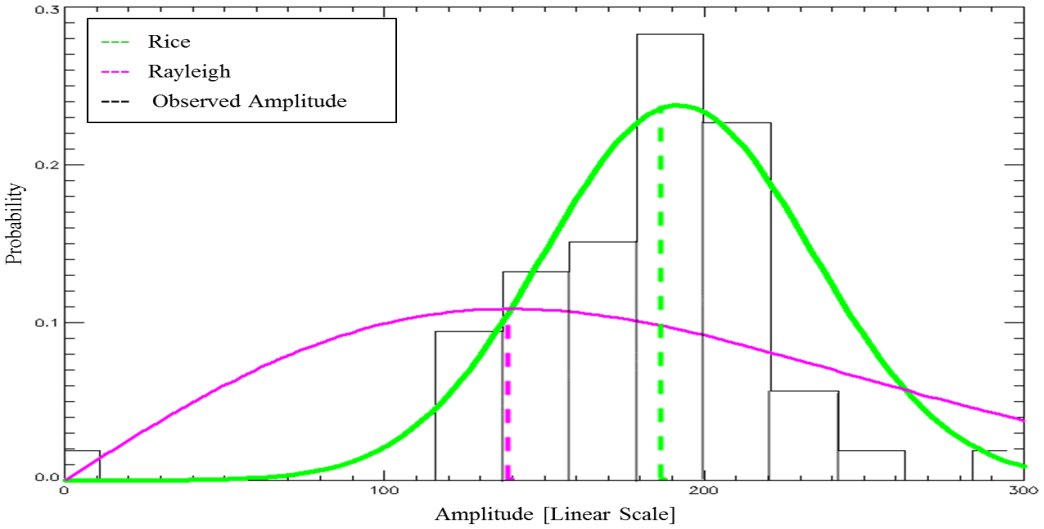



**Figure 3.9: Model fit to the amplitude stack of sample 1; the normalized histogram of the amplitude stack is overlaid with the PDF of two resulted models (scaled by factor of 20), the estimated expected signal amplitude of each of the models is shown with dashed lines. This sample scattering is perfectly described by the assumed Rice model with a high SCR level and thus is inferred to be a persistent scatterer.**

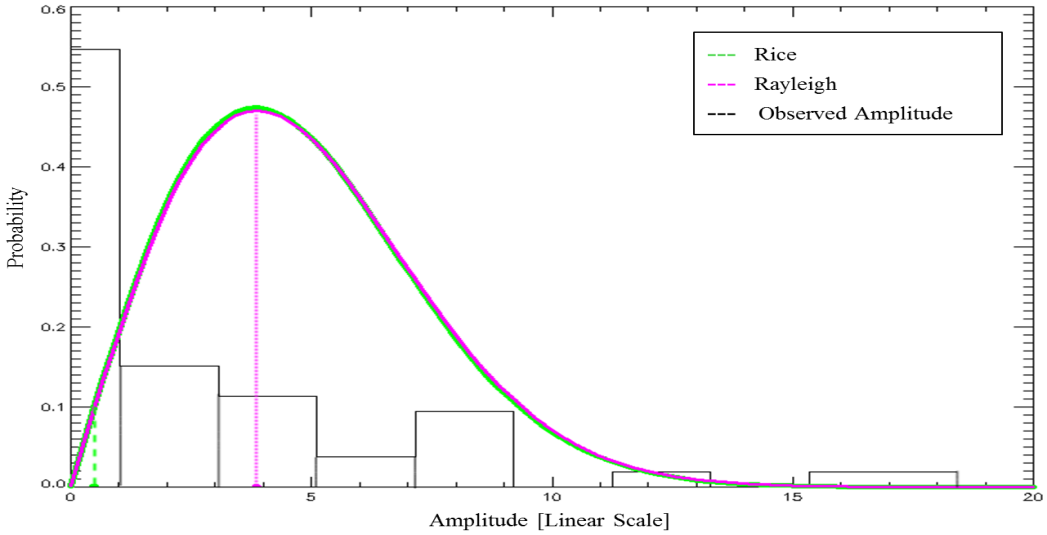



**Figure 3.10: Model fit to the amplitude stack of sample 2; the normalized histogram of the amplitude stack is overlaid with the two resulted models, the estimated expected signal amplitude of each of the models is shown with dashed lines. This example implies that the two assumed models of Rice and Rayleigh do not suffice for inference on scattering mechanisms involved in the corresponding resolution cell.**

The first sample's amplitude normalized histogram overlaid with the two fitted models is illustrated in Fig. 3.9. For visualization purposes, probability values of the two fitted PDFs are scaled by factor of 20. The estimated signal is indicated by dashed lines in each of the models. As expected the Rice model is a better fit to the histogram and the

estimated expected amplitude matches the peak of the histogram; implying a reasonable model selection and estimation result, respectively.

In contrast to the first sample, inspection of the second sample proves an inappropriate fit of the models. As shown in Fig. 3.10, the shape of the two functions are identical and none follows the data histogram; even though the Rice model is able to provide a better estimation of the expected amplitude compared to the Rayleigh. The reason is that the two considered extreme scattering cases do not suffice for description of complex scenarios. Hence more complicated probabilistic models are required in such cases, a thorough review of which can be found in [38, 39]. Later in the next section one further possibility of such scattering mechanisms is also introduced. Although poor in describing the behavior of these type of scatterers, the result of the model fit does not lead to miss detection since the estimated low SCR value prevents the point to be detected as a PS. As a conclusion, the PS detection is not compromised by this result.

The final step in detection of persistent scatterers is to apply a threshold on the SCR map. Having set the threshold to 2; equivalent to phase noise of 0.5 radians according to Eq. 3.7, the final detected PS candidates are obtained. Fig. 3.11 represents the detected PSs applying three methods of amplitude dispersion index, spatial and temporal SCR estimation, in all of which the threshold is set to $SCR = 2$. Note that the PSs detected by spatial SCR estimation method obtained from the PSI-GENESIS are much lower in number but much higher in distribution; this is due to the fact that the PSI-GENESIS software is adapted to concentrate the neighboring PSs on a single point. Therefore as opposed to the visual results, the number of PSs in this method is much more than the other two.

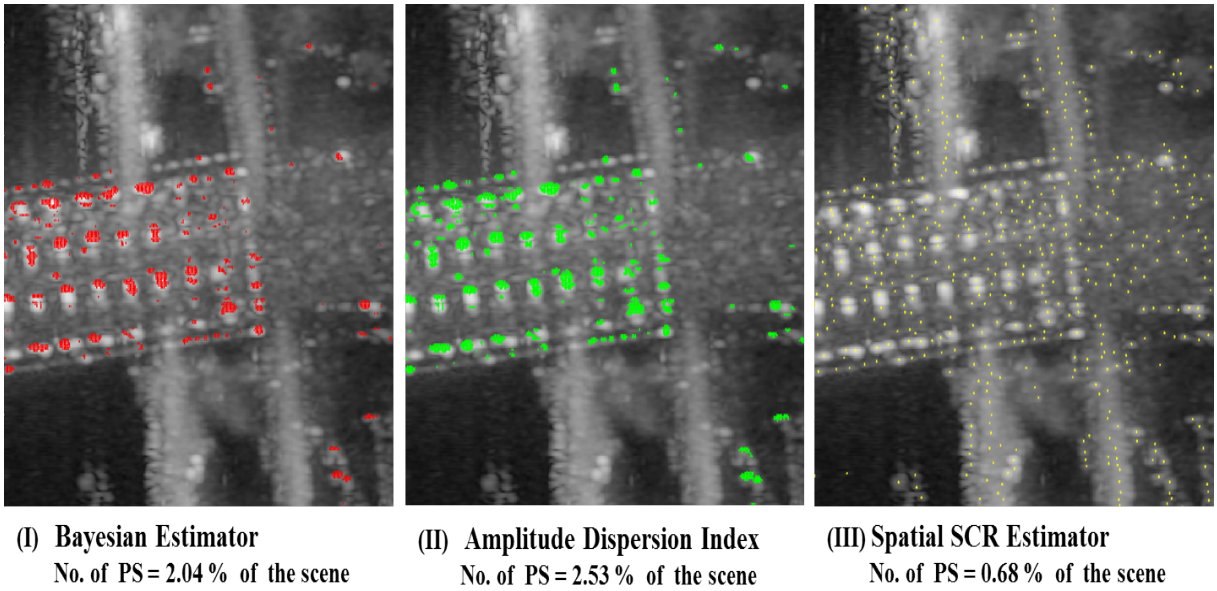



**Figure 3.11: PS detection results, applying three methods of (I) Bayesian estimation, (II) amplitude dispersion index, (III) spatial SCR estimation obtained from PSI-GENESIS. The temporal estimator is more conservative in estimation of phase error**

Comparing the results of these three approaches, it is deducted that the temporal SCR method is more conservative than its counterparts. This result could be expected beforehand since the other two methods are known to underestimate the phase error and thus detect more PSs [30].

## 3.4 Inference on Temporal Scatterers

The introduced concepts in the previous section are employed here to perform an inference on Temporal Persistent Scatterers (TPS). TPSs are scatterers with partial phase stability in a subset of interferometric stack. Temporal presence of PSs in the data stack may be resulted from changes in the acquisition geometry or physical changes occurred in the resolution cell. Construction or demolition of man-made structures as well as partial snow coverage of point scatterers are common examples of such physical changes [35].

As discussed in sub-section 3.1.1, superposition of various signals in the interferometric phase complicates the interpretation of such abrupt changes in the phase time series. In here too, alternative way is to exploit the amplitude information to detect the presence and specify the point of occurrence of such scatterers. The regular PSI analysis for extraction of signal of interest can then be performed on the time interval where the resolution cell acts as a PS.

In a stack of temporally ordered SLCs, the abrupt changes in amplitude level are indicatives for occurrence of TPS. Therefore, the TPS analysis is reformulated as detection of presence of such change points and estimation of their location as a step function. A number of methods have been proposed so far to handle the change point estimation, a comprehensive review of which is found in [40]. The focus here is on performing the change point estimation via the developed Bayesian inference software and to evaluate the performance of this approach.

The section starts with a general description of change point estimation in case of TPSs, followed by a feasibility study of the Bayesian method in this context via simulations and is finalized with the application of the method on TerraSAR-X data.

### 3.4.1 Problem Statement and Solution with Bayesian Inference

The problem is formulated as amplitude time series analysis on the stack of coregistered SLCs to detect occurrence of the TPSs and estimate the change points where the scattering characteristics changes from a PS to DS or vice verse. This goal is achieved in the following processing levels:

1. Initial classification of the scatterers into Rice and Rayleigh
2. TPS detection by performing model selection on the Rician classified pixels
3. Change point estimation for detected TPSs

Each level is further elaborated in the following:

***Initial Classification***

In order to prevent an extensive search in the entire image, subsets of pixels which are improbable to behave as a TPS are discarded in the first level. The model selection between the Rice and Rayleigh distributions is chosen here as the key for filtering such pixels.

From experiences with real data, it is observed that in a model selection between Rice and Rayleigh and in presence of a significant signal in the time series, even in a small subset of SLCs, the amplitude behavior tends toward the Rice regime. In such cases, the significant signal amplitude ($\nu$) is estimated to be low and the signal variations in the amplitude stack is interpreted as high clutter amplitude ($\sigma$) (Eq. 3.14). The corresponding SCR value is therefore estimated to be very low (this fact is also observable from the experiments of sub-section 3.3.2). Consequently, it is fair to assume that there is no possibility to find TPSs where the amplitude stack is better described by Rayleigh and finally the TPS analysis is restricted to Rice classified pixels.

It is worth mentioning that this initial model selection does not impose additional computational burden since the resulted evidence in this level is further used in the TPS detection level as well.

The alternatives for this initial classification might be the spatial SCR estimator or the amplitude dispersion index. Although the decision on the detection threshold is critical for both mentioned methods and thus the proposed model selection between Rice and Rayleigh is preferred.

### *TPS Detection*

In the current step, the Rice classified pixels are tested to detect the TPSs. The behavior of TPS in amplitude stack is narrowed down to two different scattering cases; fading and appearance of PSs on a background clutter. As discussed before, the scattering mechanism of PS and DS is described by Rice and Rayleigh distribution, respectively. Therefore the TPSs are described by conversion from one scattering regime to the other (Fig. 3.12).

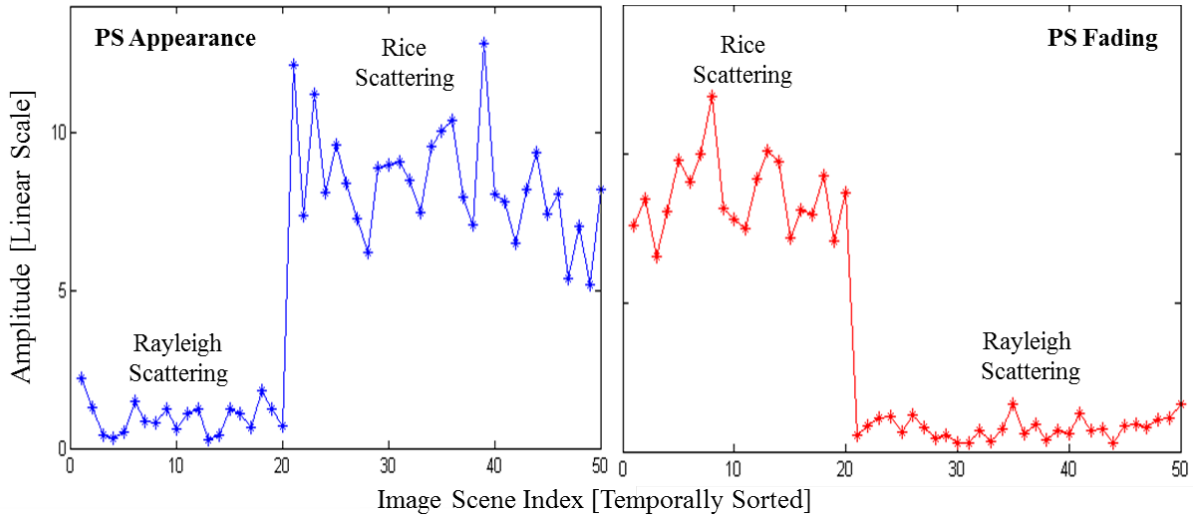



**Figure 3.12: Simulation of amplitude stack showing the two possible scenarios for temporal persistent scatterers; The TPS narrows down to appearance or fading of the persistent scatterers with its change point described by the combination of two scattering functions**

The detection of TPSs is thus based on a Bayesian model selection between pure Rice scattering and a combination of Rice and Rayleigh scattering model.

Considering a single change point in the TPS, the model selection is performed by segmentation of the amplitude stack into two parts and fitting the Rice-Rayleigh or Rayleigh-Rice combined models to the paired segments. The Bayesian networks of these two cases are provided in Fig. 3.13. The process is mathematically summarized in the following steps:

$$A = [a_1, a_2, \ldots, a_N] \Longrightarrow s_1 = [a_1, \ldots, a_{N/2}], s_2 = [a_{(N/2)+1}, \ldots, a_N]$$

$$E_{Initial} = p(\nu, \sigma | A, M_{Rice})$$

$$E_{Fading} = p(\nu, \sigma | s_1, M_{Rice}) \times p(\sigma | s_2, M_{Rayleigh})$$

$$E_{Appearance} = p(\sigma | s_1, M_{Rayleigh}) \times p(\nu, \sigma | s_2, M_{Rice})$$

$$E_{max} = \text{Max}([E_{Initial}, E_{Fading}, E_{Appearance}])$$

If $E_{max} = E_{Initial} \Longrightarrow$ No TPS is detected

If $E_{max} = E_{Fading} \Longrightarrow$ Fading TPS is detected

If $E_{max} = E_{Appearance} \Longrightarrow$ Appearing TPS is detected

Where $E$ denotes the evidence and $M$ represents the model. In case a TPS is detected the maximum evidence between the fading and appearing models give an initial hint on the type of TPS, namely fading or appearing scenario, to be further regarded in change point estimation.

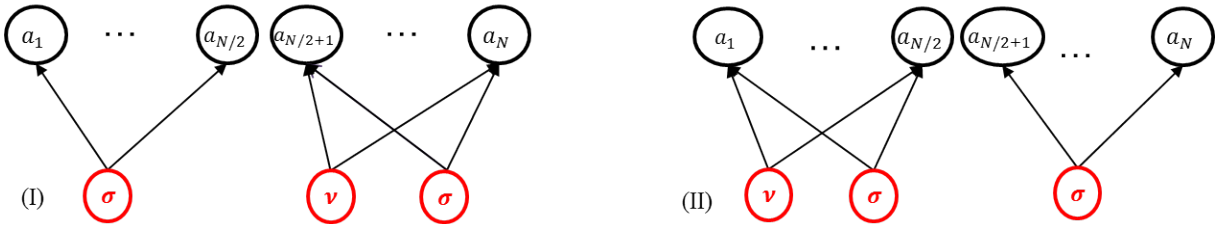



**Figure 3.13: Bayesian network of combined scattering scenarios for modeling the (I) PS appearance and (II) PS fading in the amplitude stack**

### *Change Point Estimation*

So far the TPS is detected and its generic type as an appearing or fading PS is specified. Final step is to estimate the exact change point in which the conversion between the scattering model occurs. The model selection is used here as well to perform the estimation; the amplitude stack is segmented at different indexes of the stack and the specified combined scattering model is fitted to the pairs. Finally the change point is estimated as the segmentation point $m$ that gives the highest evidence:

$$A = [a_1, a_2, \ldots, a_N] \longrightarrow s_{m,1} = [a_1, \ldots, a_m] \qquad s_{m,2} = [a_{m+1}, \ldots, a_N]$$
$$\hat{m} = \text{argmax}_m \{p(x_1|s_{m,1}, M_1) \times p(x_2|s_{m,2}, M_2)\}$$

In here, $x_i$ represents the vector containing the parameters of the model $M_i$, and models are chosen between the Rice and Rayleigh according to the type of TPS specified in the TPS detection step. To further decrease the computational burden, the search for change point is performed in a coarse and fine step.

### 3.4.2 Investigation with Synthetic Data

The proposed method is evaluated by simulation of stack of 50 amplitude observations imitating the fading TPS scenario. The clutter and signal power in Rice and Rayleigh segments of the stack are kept fixed and variation of change point as well as the SCR of the TPS is studied in the following.

In the first case, SCR of the Rice segment is varied in the range of 1 to 10 while the change point is kept fixed and in the middle of the time series. The trial is repeated 50 times and the probability of missed detection of the TPS as well as success rate in correct estimation of change point is reported. Fig. 3.14 summarizes the obtained results.

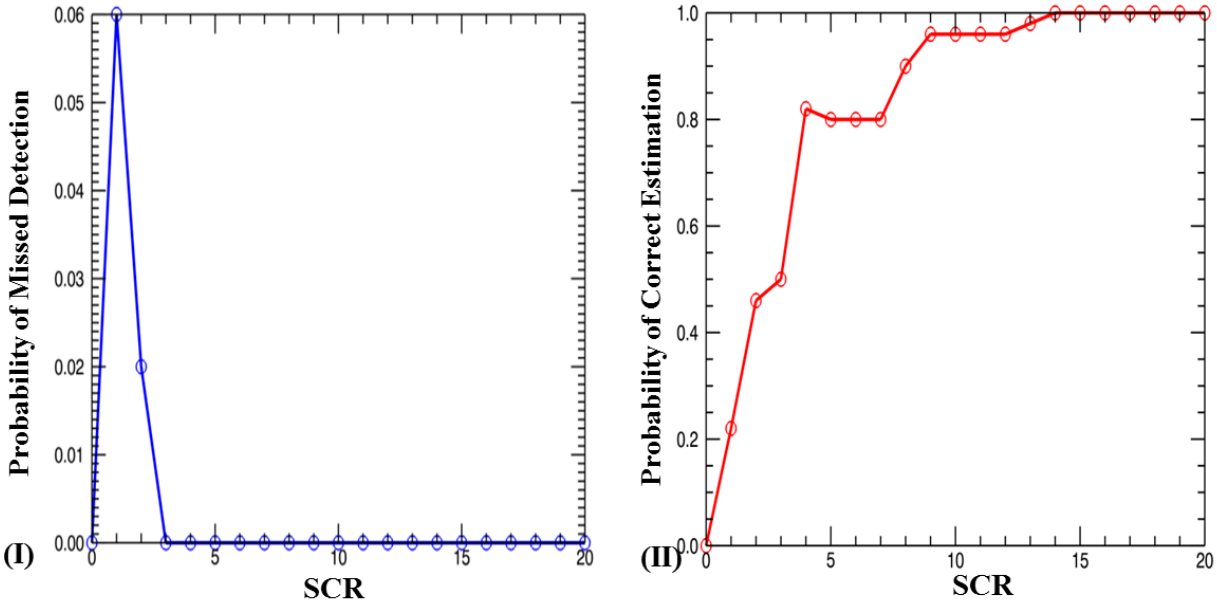



**Figure 3.14: TPS detection and change point estimation results; impact of SCR variation on (I) the missed detection rate of the TPS and (II) correct estimation of the change point**

As it is evident, the applied method is quite robust in detecting the presence and type of the TPS; since the missed detection probability is extremely low for SCR level of less than 3 and completely improbable for higher SCR levels.

In the second case the impact of number of observations in the model fit is assessed. For this purpose, the change point is simulated between the SLC indexes of 1 to 49 in steps of 2 points while the SCR is set fixed to 5. The trial is repeated 50 times and the probabilities of missed detection together with the correct estimation rate are reported (Fig. 3.15). Several points can be deduced from the results:

1. As expected, the number of observations has a direct impact on the performance. The change points occurring in the middle of the stack are easier to detect and estimate since nearly equal number of observations are provided for each of the Rice and Rayleigh fitting.
2. Bearing in mind that the Rice-Rayleigh combination was the simulated scenario, it is seen that the number of observations is more critical in Rice model fit since the change points in the beginning of the stack are more difficult to detect. This fact is related to the complexity of the Rice model compared to Rayleigh.
3. At least 10 observations are required to have an acceptable performance.
4. Comparing the current case with SCR variation results, it is evident that performance of the method is more affected by the change point position compared to the SCR.

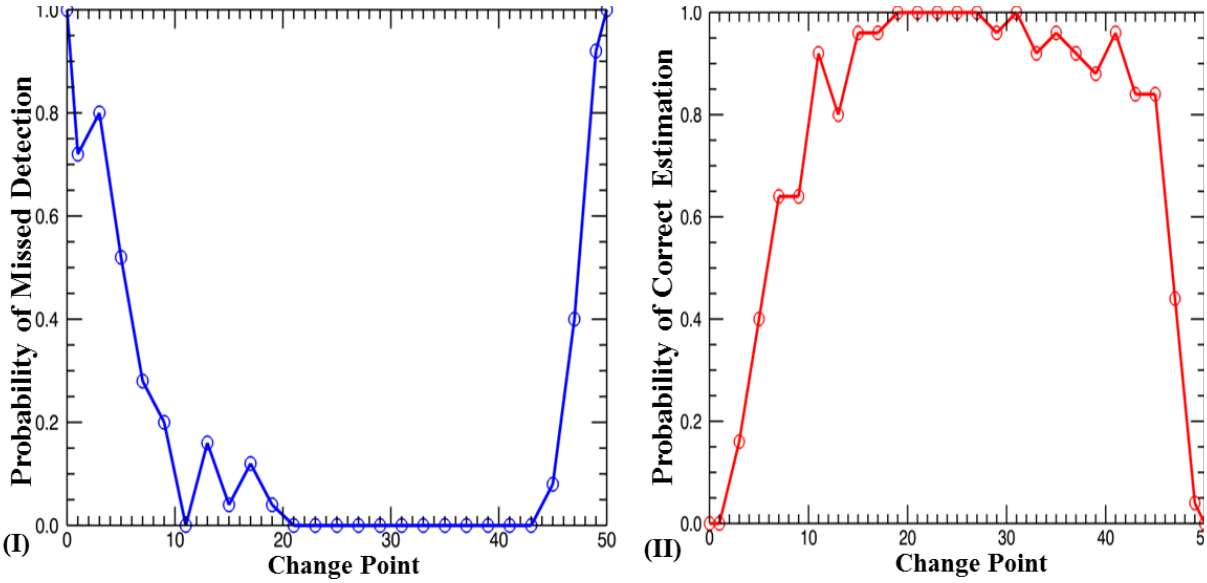



**Figure 3.15: TPS detection and change point estimation results; impact of change point variation on (I) the missed detection rate of the TPS and (II) correct estimation of the change point**

At last but not least, the computational cost of the approach has also been investigated through the simulations. Results from all the simulated cases reveal that the detection of TPSs and estimation of their change point in the three mentioned levels in sub-section 3.4.1, takes in average 7 seconds for stack of 50 observations. This result is obtained using a double-core OptiPlex 740 machine with 8 GB of system memory and with the IDL language.

To have a comprehensive evaluation, the method is further applied in analysis of real SAR data stacks in the next section.

### 3.4.3 Experiment on Real Data

To justify the importance of TPS detection, a special test site is chosen in this section for which the TPS analysis is the key method in revealing the deforming signal. The test site corresponds to the recently constructed *Federal Intelligence Service (BND)* building in Berlin, Germany (Fig. 3.16). Due to partial phase stability of the scatterers on the building structure, the building is completely ignored by the coherent InSAR techniques.

**Figure 3.16: Test site 2, optical image of the new Federal Intelligence Service building, Berlin-Germany November 2012 (Source: Google Earth)**

A stack of 79 spotlight mode TerraSAR-X SLCs are acquired over this area. The acquisition time span is from February 2008 to February 2012, while construction of the building started around 2009 and still continues to date (2013). Fig. 3.17 shows the estimated deformation map of the Berlin city processed by PSI-GENESIS, with the BND building shown to be uncovered by the PSI analysis.

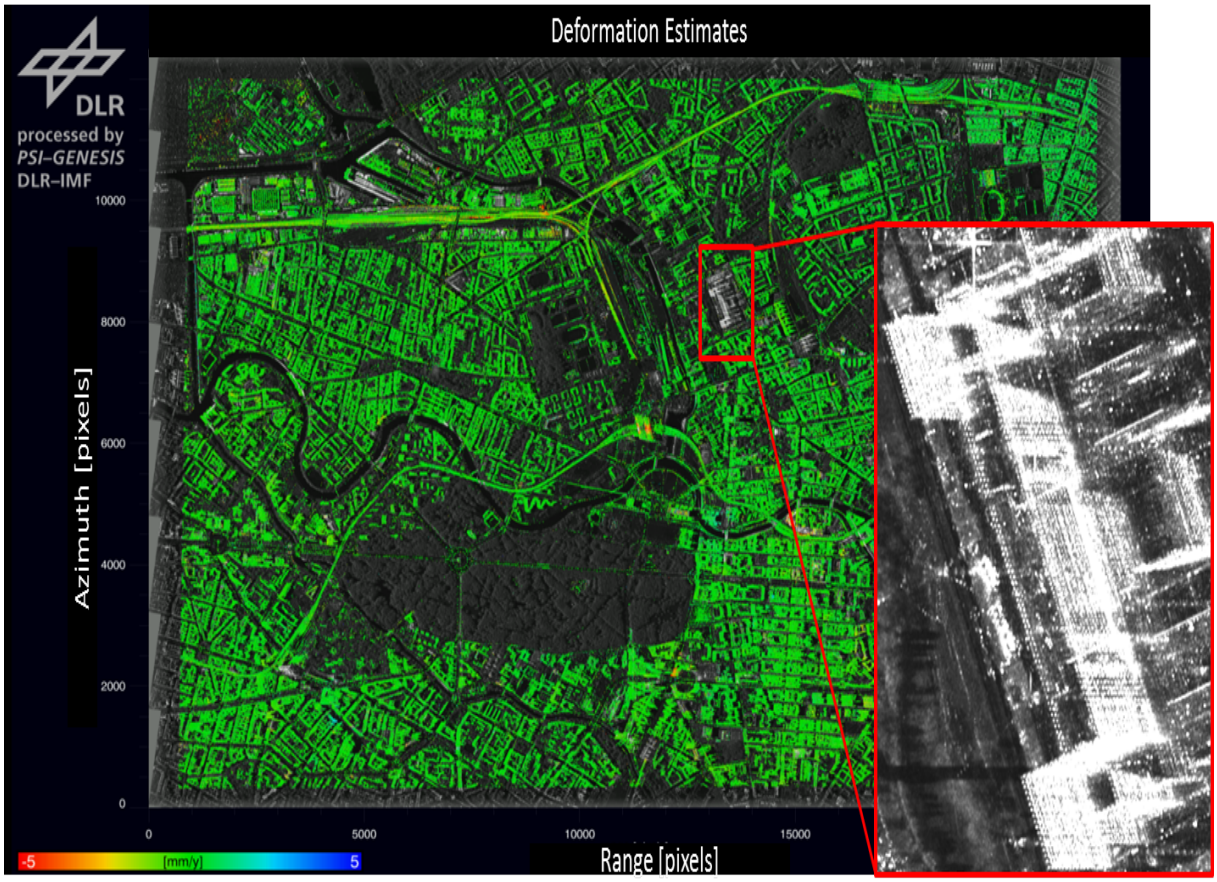
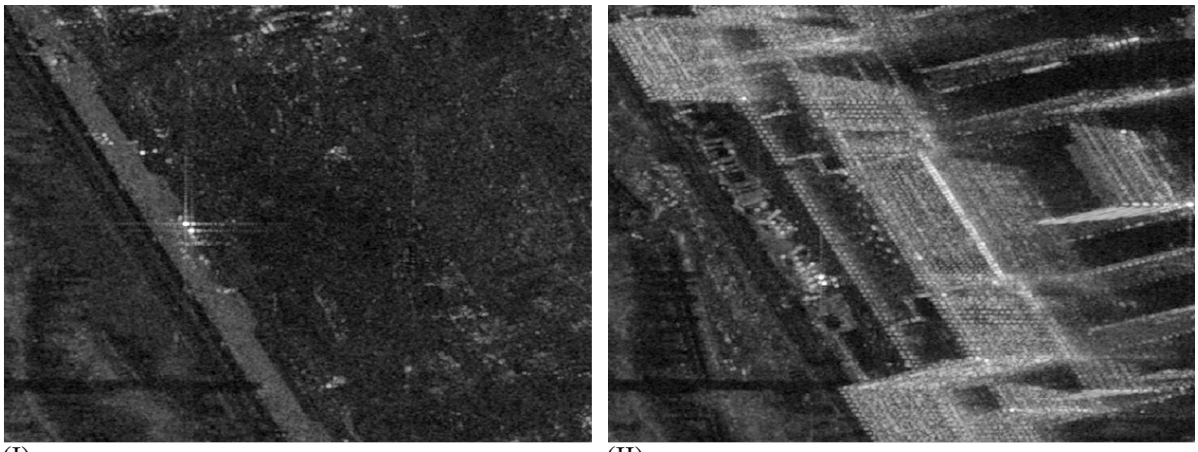



**Figure 3.17: Deformation map of Berlin city estimated by PSI-GENESIS with the BND building uncovered in the deformation map; being appeared in the middle of the acquisition period, the scatterers corresponding to the BND building are all TPSs and therefore are not included in the PSI time series analysis**



(I) (II)

**Figure 3.18: Calibrated multi-looked amplitude of the test site at (I) 10 first SLCs of the stack acquired in 2008 and (II) 10 last SLCs acquired in 2011-2012**

Applying the TPS detection algorithm on the test site the classification result is obtained. Fig. 3.19 illustrates the resulted classification map with the three distinguished scattering type of DS, PS and TPS. As it is seen most of the pixels on the building are detected as "appearing" TPSs. A glance at a first and last acquisition amplitude image provided in Fig. 3.18 validates the detection result. Since in the first acquisitions the test site is a bare land and in the last ones appearance of the building is evident.

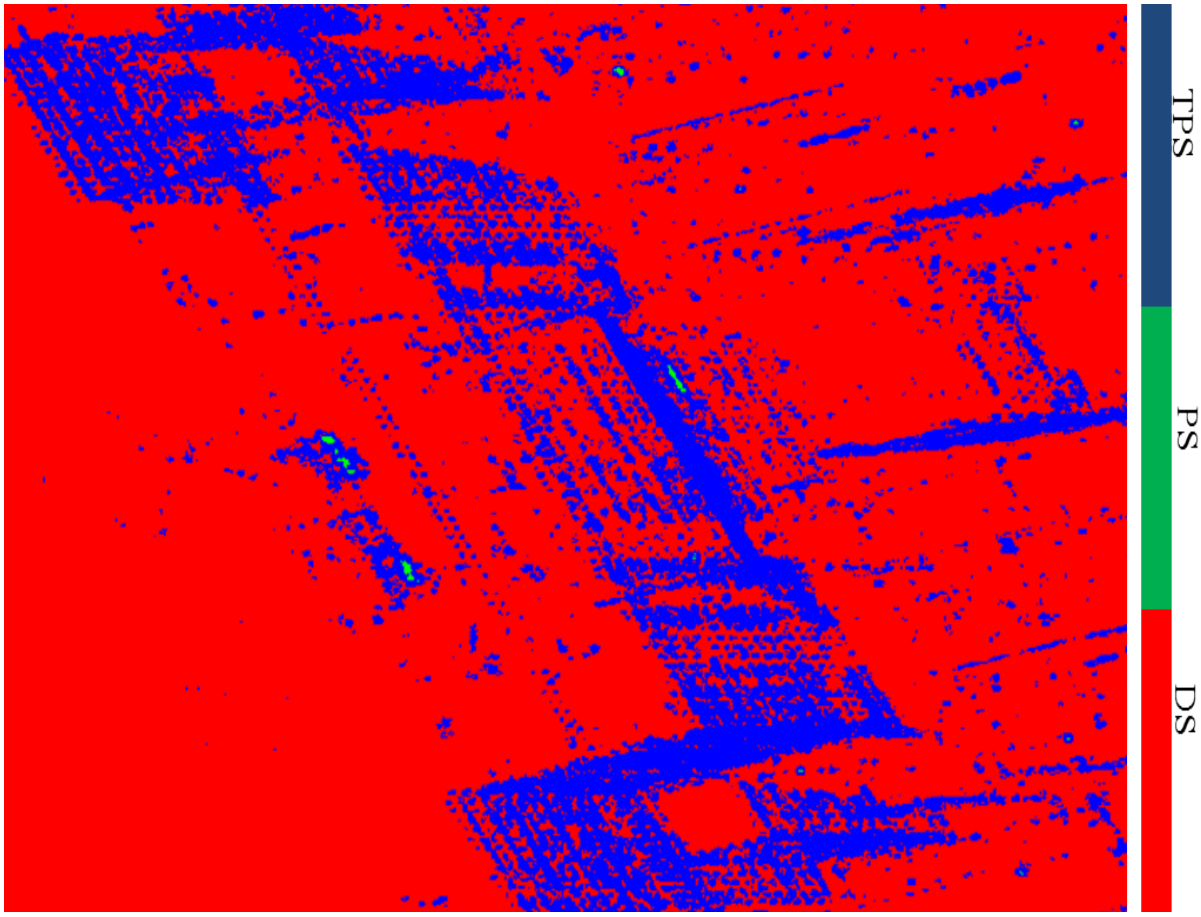



**Figure 3.19: Pixel classification resulted from TPS detection; three types of scatterers are distinguished: the temporal persistent scatterers coded in blue, the persistent scatterers coded in green and the distributed scatterers, coded in red color. 85.2 percent of the scene is detected to be DS, less than one percent PS and 14.5 percent TPS.**

The change point estimation is considered for the detected TPS cells. The change points are color coded in Fig. 3.20. In this specific test site the change points indicate the construction time of each part of the building. The estimation result provided in Fig. 3.20 shows the building construction progress starting from 2009 to 2012.

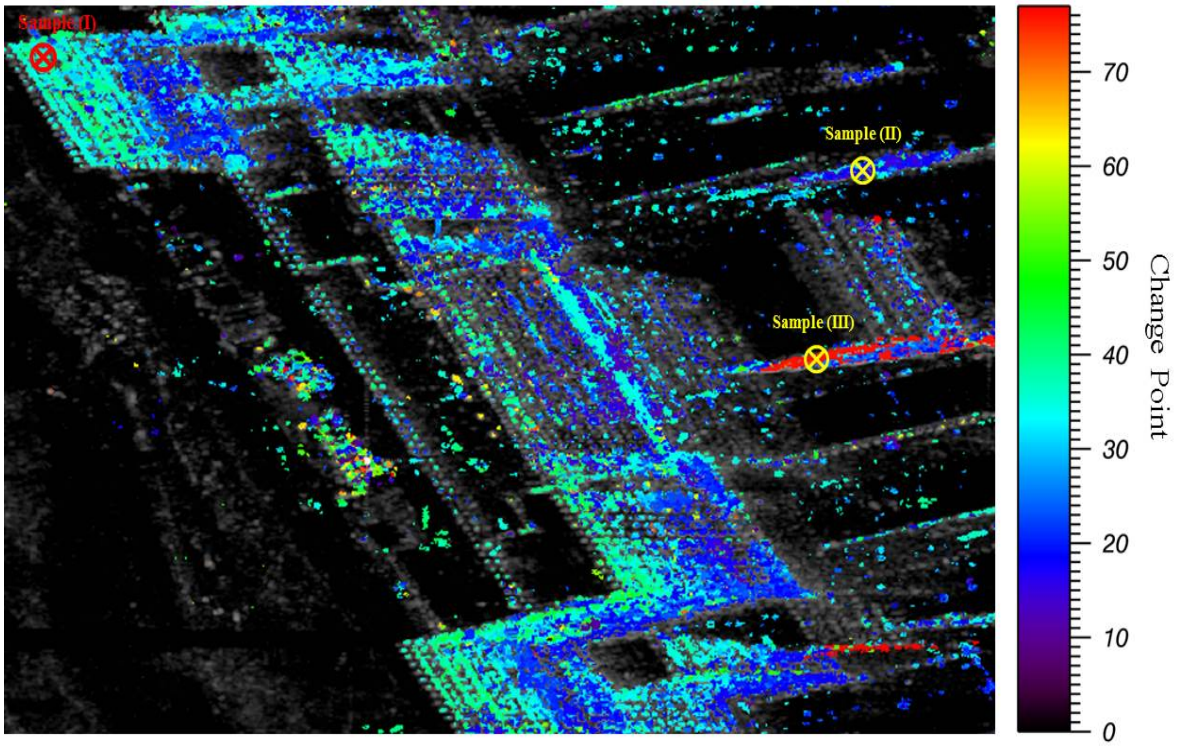



**Figure 3.20: Result of change point estimation on the detected TPSs; the change point implies the construction progress of each part of the building starting from 2009 and continuing until 2012. the marked samples are further inspected.**

The detected TPSs with respect to the time of acquisition is summarized in the Fig. 3.21, the number of TPSs are reported in percentage. The plot implies the start of the building construction at around 2009 which is aligned with the reported progress of the building.

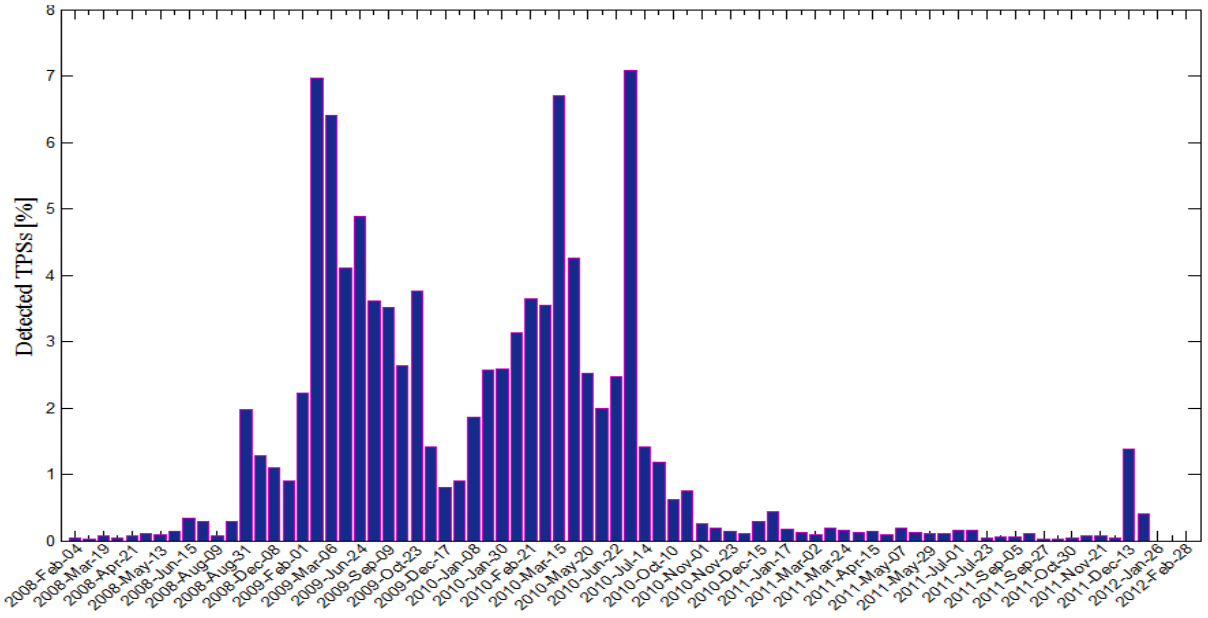



**Figure 3.21: Percentage of detected TPSs versus acquisition time of the SAR images; the bars imply the construction progress of the building over time**

Specified by crosses in Fig. 3.20, three samples are chosen in the scene to be further inspected. The amplitude time series of these samples together with their estimated change points are depicted in Fig. 3.22. The visualized results reveal that the change point estimation is successful, even in challenging cases such as the third sample where the change points occurred towards the end of stack leaving only 3 amplitude observations in the segment for fitting the Rician model.

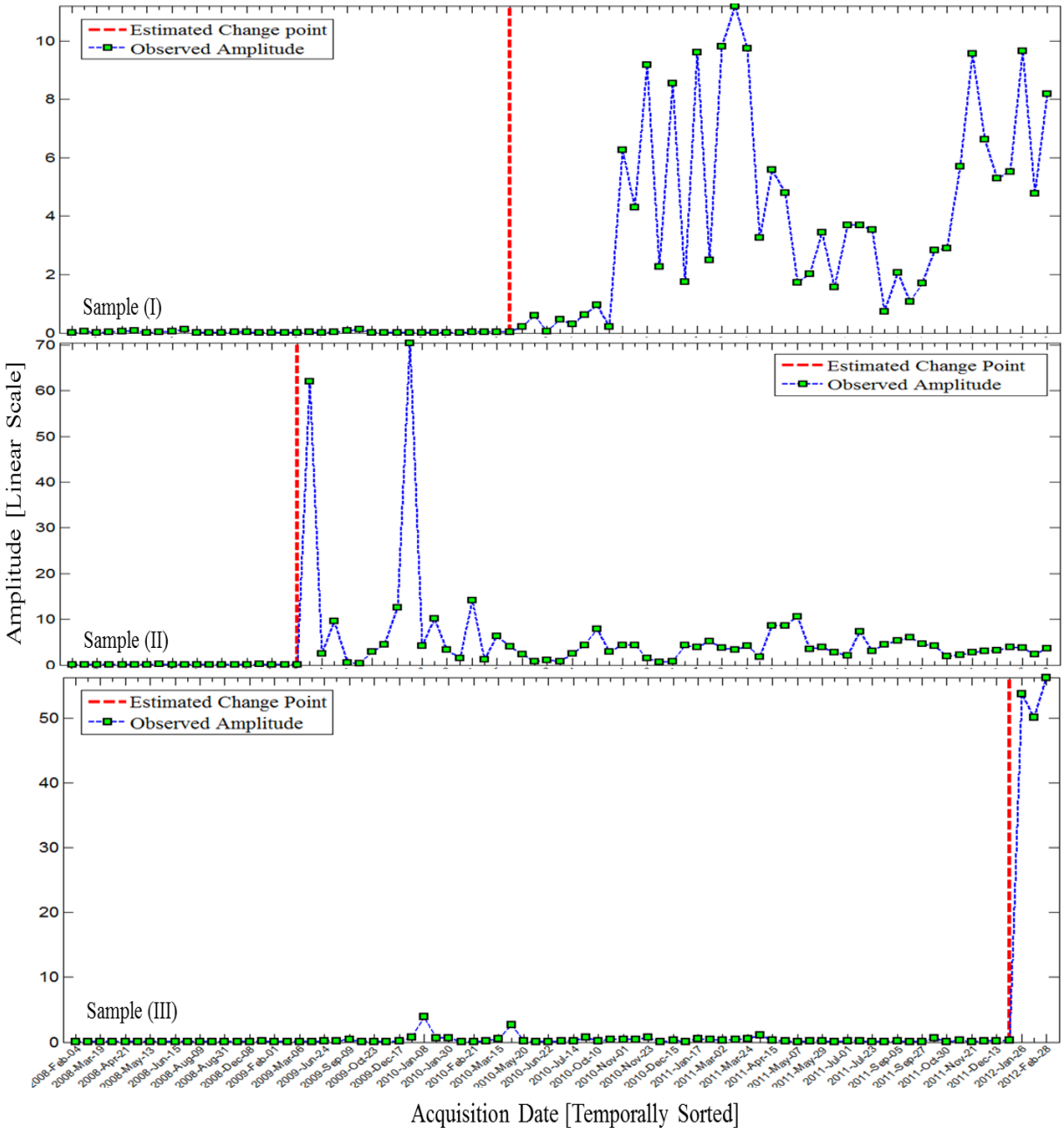



**Figure 3.22: Change point estimation results of the three TPS samples specified in Fig. 3.20; the software performance is satisfactory even in estimation of challenging change points occurring at the beginning or the end of the amplitude time series**

# Chapter 4

# Concluding Remarks

The current thesis presents the investigation of SAR amplitude time series and its relevance to temporal phase coherence as assistance to advanced InSAR techniques such as PSI and differential TomoSAR. The involved inference tasks of model selection and estimation has been carried out using a developed tool based on the Bayesian statistics. The developed software deals with inverse problems in a well established probabilistic framework and is adaptable to any defined probabilistic model.

The highlights of the thesis as well as the areas of further improvements are summarized in the following sections.

## 4.1 Conclusion and Discussion

Having performed amplitude time series analysis on the simulated as well as the real data, the following concluding remarks are drawn:

- The amplitude of the SAR complex valued images provides the potential to investigate the sub-resolution scattering phenomena.
- Phase coherence is the first and foremost criterion for the InSAR techniques. The phase noise estimation is possible through estimation of the signal parameters from a considered point scatterer model. Taking this approach, the estimation bias of the available amplitude-based methods is reduced.
- In the advanced techniques the constraint on the long time phase coherence might be too strict and lead to loss of information in the final results. Partial phase coherence of the InSAR stacks can be assessed by considering a combination of scattering models. The temporal persistent scatterers can be detected and integrated in the advanced InSAR techniques in order to improve the information content of their final product.
- The potential of Monte-Carlo based approaches in the solution of Bayesian inference is proved in the analysis. The analytically intractable or complicated inverse problems can be formulated and handled via the developed Monte-Carlo-based inference framework. Compared to the alternative methods, the Monte-Carlo approach is less

complicated in terms of mathematical manipulation and more accurate in terms of performance.

## 4.2 Outlook

The following possibilities are suggested for future extension of the current work:

- Consideration of K-distribution in the scattering model for further classification of distributed scatterers
- Improvement of the scattering model to infer the number of dominant scatterers in the resolution cell
- Consideration of multiple change points in the analysis of the temporal persistent scatterers
- Consideration of numerically stable algorithms in execution of arithmetic manipulation in the Bayesian network to improve the accuracy of evidence calculation in the model selection
- Integration of outlier detection methods to the Bayesian network in order to minimize the impact of the falsifying observed random variables in the inference

Apart from the aforementioned potential improvements in the content of this work, the developed Bayesian inference tool opens new opportunities for further applications in the InSAR techniques. An interesting application could be the model selection between the relevant deformation models and estimation of nonlinear deformation parameters in the InSAR stacks.

# Appendix A

**Overview of the developed Bayesian inference software:** The following tables provide an overview of the modules of the developed software. The modules are listed in the order they are called by the main routine.

| Module | Called by Superior Module | Input | Output |
|---|---|---|---|
| **1.Main** | ... | 1.Information on Nodes and PDFs | 1.DGM<br>2.Evidence<br>3.Estimated Latent RVs |
| **2.Bayesian Network** | Main | Information on Nodes and PDFs | DGM |
| **3.Learning** | Main | 1.DGM<br>2.Number of Samples | 1.Evidence<br>2.Important Area of the Latent RVs<br>3.Optimum Number of Samples for each Latent RV |
| **4.Samples From PDF** | 1.Learning<br>2.Importance Sampling | DGM | 1.Generated Particles<br>2.Particles' Weights (as Prior Probability) |
| **5.Evaluate PDF** | 1.Learning<br>2.Importance Sampling | DGM | Likelihood Probability for the Particles |
| **6.Inference** | 1.Learning<br>2.Importance Sampling | 1.Set of samples of Latent RV<br>2.Non-normalized Posterior PDF<br>3.Flag Specifying Functionality of the Module "MAP", "Evidence" | Evidence or MAP Estimate |

| Module | Called by Superior Module | Input | Output |
|---|---|---|---|
| **7.Importance Area** | Learning | 1.Set of samples of Latent RV<br>2.Normalized Posterior PDF<br>3.Index of Target Latent RV | 1."Important" Area of the Prior PDF for the Target Latent RV<br>2.Optimum Number of Samples for the Target RV |
| **8.Importance Resampling** | Main | 1.DGM<br>2.Optimum Number of Samples<br>3.Maximum Allowed Number of Samples<br>4.Minimum Allowed Number of Samples<br>5.Minimum allowed number of samples<br>6."Important" Area of the Latent RVs | 1.Estimated Value for the Latent RVs<br>2.Improved Evidence after Importance Resampling |
| **9.PDF Estimation** | Importance Resampling | 1.Set of Samples of Latent RVs<br>2.Normalized Posterior PDF<br>3.Type of Kernel<br>4.Index of Target Latent RV<br>5.Number of Points on the Regular Grid | 1.Regular Grid on Target Latent RV<br>2.Estimated PDF |
| **10.Marginalization** | 1.Importance Area<br>2.PDF Estimation | 1.Set of Samples of Latent RVs<br>2.Normalized Posterior PDF<br>3.Index of Target Latent RV | Marginalized PDF of Target Latent RV |
| **11.CDF** | 1.Importance Area<br>2.PDF Estimation | 1.Samples of Target Latent RV<br>2.Marginalized PDF | CDF of Target Latent RV |
| **12.Resampling Systematic** | 1.Importance Area<br>2.PDF Estimation | 1.Sorted Particles<br>2.CDF<br>3.Number of Resampled Particles | Resampled Particles |
| **13.Kernel Estimator** | PDF Estimation | 1.Resampled Particles<br>2.Kernel Type | 1.Regular Grid on Target Latent RV<br>2.Estimated PDF |

# Bibliography


[1] Richard Bamler and Philipp Hartl. Synthetic aperture radar interferometry. *Inverse problems*, 14(4):R1, 1998.

[2] Ramon F Hanssen. *Radar interferometry: data interpretation and error analysis*, volume 2. Springer, 2001.

[3] Alessandro Ferretti, Claudio Prati, and Fabio Rocca. Permanent scatterers in sar interferometry. *Geoscience and Remote Sensing, IEEE Transactions on*, 39(1):8–20, 2001.

[4] Andreas Reigber and Alberto Moreira. First demonstration of airborne sar tomography using multibaseline l-band data. *Geoscience and Remote Sensing, IEEE Transactions on*, 38(5):2142–2152, 2000.

[5] Xiao Xiang Zhu and Richard Bamler. Very high resolution spaceborne sar tomography in urban environment. *Geoscience and Remote Sensing, IEEE Transactions on*, 48(12):4296–4308, 2010.

[6] Howard A Zebker and John Villasenor. Decorrelation in interferometric radar echoes. *Geoscience and Remote Sensing, IEEE Transactions on*, 30(5):950–959, 1992.

[7] Devinderjit Sivia and John Skilling. *Data analysis: a Bayesian tutorial*. Oxford University Press, USA, 2006.

[8] Christopher M Bishop. *Pattern recognition and machine learning*, volume 1. springer New York, 2006.

[9] PJG Teunissen, DG Simons, and CCJM Tiberius. Probability and observation theory. *Lecture Notes Delft University of Technology*, 2005.

[10] Joseph JK O'Ruanaidh and William J Fitz Gerald. *Numerical Bayesian methods applied to signal processing*, volume 5. Springer-Verlag New York, 1996.

[11] Eric W. Weisstein. Monte carlo method. from MathWorld – A Wolfram Web Resource, Visited on: 30.09.2013. http://mathworld.wolfram.com/MonteCarloMethod.html.

[12] Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. An introduction to mcmc for machine learning. *Machine learning*, 50(1-2):5–43, 2003.

[13] Jun S Liu. *Monte Carlo strategies in scientific computing*. springer, 2008.

[14] Walter R Gilks, Sylvia Richardson, and David J Spiegelhalter. *Markov chain Monte Carlo in practice*, volume 2. CRC press, 1996.

[15] M Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *Signal Processing, IEEE Transactions on*, 50(2):174–188, 2002.

[16] Jeroen D Hol, Thomas B Schon, and Fredrik Gustafsson. On resampling algorithms for particle filters. In *Nonlinear Statistical Signal Processing Workshop, 2006 IEEE*, pages 79–82. IEEE, 2006.

[17] Y Boers. On the number of samples to be drawn in particle filtering. In *Target Tracking: Algorithms and Applications (Ref. No. 1999/090, 1999/215), IEE Colloquium on*, pages 5–1. IET, 1999.

[18] M Vidyasagar. Statistical learning theory and randomized algorithms for control. *Control Systems, IEEE*, 18(6):69–85, 1998.

[19] Bruce E Hansen. Lecture notes on nonparametrics. *Lecture notes University of Wisconsin*, 2009.

[20] Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, 6 (2):461–464, 1978.

[21] Ritei Shibata. *Statistical aspects of model selection*. Springer, 1989.

[22] Hamparsum Bozdogan. Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, 52(3):345–370, 1987.

[23] Peter Grünwald, Petri Kontkanen, Petri Myllymäki, Tomi Silander, and Henry Tirri. Minimum encoding approaches for predictive modeling. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 183–192. Morgan Kaufmann Publishers Inc., 1998.

[24] Peter D Grünwald, In Jae Myung, and Mark A Pitt. *Advances in minimum description length: Theory and applications*. MIT press, 2005.

[25] Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995.

[26] PT Troughton. Simulation methods for linear and nonlinear time series models with application to distorted audio signals. *Unpublished Ph. D thesis, University of Cambridge*, 1999.

[27] James O Berger and Luis R Pericchi. The intrinsic bayes factor for model selection and prediction. *Journal of the American Statistical Association*, 91(433):109–122, 1996.

[28] William B Bishop and Petar M Djuric. Model order selection of damped sinusoids in noise by predictive densities. *Signal Processing, IEEE Transactions on*, 44(3): 611–619, 1996.

[29] Bert M Kampes. *Radar interferometry: persistent scatterer technique*, volume 12. Springer, 2006.

[30] Nico Adam, Bert Kampes, and Michael Eineder. Development of a scientific permanent scatterer system: Modifications for mixed ers/envisat time series. In *Proceedings of the 2004 Envisat & ERS Symposium*, 2004.

[31] Franz Meyer, Stefan Gernhardt, and Nico Adam. Long-term and seasonal subidence rates in urban areas from persistent scatterer interferometry. In *Urban Remote Sensing Joint Event, 2007*, pages 1–6. IEEE, 2007.

[32] Nico Adam, Alessandro Parizzi, Michael Eineder, and Michele Crosetto. Practical persistent scatterer processing validation in the course of the terrafirma project. *Journal of Applied Geophysics*, 69(1):59–65, 2009.

[33] ESTEC. Sar calibration workshop. In *Noordweijk, Netherlands*, Septamber 1993.

[34] Andrew John Hooper. *Persistent scatter radar interferometry for crustal deformation studies and modeling of volcanic deformation*. PhD thesis, Stanford University, 2006.

[35] Alessandro Ferretti, Carlo Colesanti, Daniele Perissin, Claudio Prati, and Fabio Rocca. Evaluating the effect of the observation time on the distribution of sar permanent scatterers. In *Proc. of FRINGE 03, ESA-ESRIN, Frascati (Italy)*, pages 1–5, 2003.

[36] Eric W. Weisstein. Modified bessel function of the first kind. from MathWorld – A Wolfram Web Resource, Visited on: 30.09.2013. http://mathworld.wolfram.com/ModifiedBesselFunctionoftheFirstKind.html.

[37] Cheng Guan Koay and Peter J Basser. Analytically exact correction scheme for signal extraction from noisy magnitude mr signals. *Journal of Magnetic Resonance*, 179(2):317–322, 2006.

[38] Gabriele Moser, Josiane Zerubia, and Sebastiano B Serpico. Sar amplitude probability density function estimation based on a generalized gaussian model. *Image Processing, IEEE Transactions on*, 15(6):1429–1442, 2006.

[39] Céline Tison, J-M Nicolas, Florence Tupin, and Henri Maître. A new statistical model for markovian classification of urban areas in high-resolution sar images. *Geoscience and Remote Sensing, IEEE Transactions on*, 42(10):2046–2057, 2004.

[40] Ramon Brcic and Nico Adam. Detecting changes in persistent scatterers. In *IEEE International Geoscience and Remote Sensing Symposium, 21.-26. Juli 2013*, pages 1–4. IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2013.